

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:25:02 ; Search time 28.15 Seconds
(without alignments)
337.411 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401
Sequence: 1 MOYDETLIPKRVPSLCSARY.....LIFGEADVQEMAKERKILRL 401

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 188963 seqs, 23686106 residues

Word size : 0

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	401	100.0	401	1 W69971	Human sodium-depen
2	241	60.1	401	1 W69920	Human haemochromat
3	9	2.2	436	1 W78919	Human haemochromat
4	7	1.7	12	1 R56478	Vitamin E transport
5	7	1.7	20	1 R15569	Immunopeptide #2 d
6	7	1.7	29	1 R31212	HPV-16 E7 peptide.
7	7	1.7	30	1 W93288	Human papillomavir
8	7	1.7	57	1 R66178	Mouse T18 Flt3 il
9	7	1.7	86	1 R72968	Pig kidney cell mu
10	7	1.7	98	1 R22767	HPV E7 peptide. Im
11	7	1.7	98	1 R42361	Human papillomavir
12	7	1.7	98	1 W46886	Amino acid sequenc
13	7	1.7	172	1 R97562	Human papilloma vi
14	7	1.7	185	1 W93370	Papillomavirus E7/
15	7	1.7	220	1 Y02634	Prot.D1/3-E7-mut(C
16	7	1.7	220	1 Y02631	Protdhrl26-E7-His
17	7	1.7	239	1 Y02636	CLYTR-E7-His prote
18	7	1.7	253	1 W81586	CTLA4/E7 fusion pr
19	7	1.7	253	1 W87562	Amino terminal CTL
20	7	1.7	253	1 Y01502	Amino terminal CTL
21	7	1.7	253	1 W97612	HPV 16 E7 protein
22	7	1.7	262	1 R27724	Human papilloma vi
23	7	1.7	266	1 R97561	Porcine mutarotase
24	7	1.7	341	1 R70142	Pig kidney cell mu
25	7	1.7	341	1 W26964	Human chemokine re
26	7	1.7	344	1 W26767	Amino acid sequenc
27	7	1.7	356	1 W23957	Human macrophage/d
28	7	1.7	356	1 W48087	Prot.D1/3-E7-His
29	7	1.7	371	1 Y02633	CLYTR-E6E7-His pro
30	7	1.7	390	1 Y02637	Human brain sodium
31	7	1.7	560	1 W05148	Human sodium-11thl
32	7	1.7	560	1 W70500	Eat-4 protein amin
33	7	1.7	576	1 W88523	

34	7	1.7	671	1 R85290	Streptococcus faec
35	6	1.5	11	1 R33150	HPV E7 protein - R
36	6	1.5	13	1 R10628	Human Papilloma VI
37	6	1.5	20	1 W40251	Human wild-type E-
38	6	1.5	20	1 W36605	Human E-Cadherin v
39	6	1.5	22	1 W06566	Human preprocollag
40	6	1.5	24	1 R33151	HPV E7 protein - R
41	6	1.5	33	1 W40008	Peptide effecting
42	6	1.5	40	1 R96006	Asymmetry sequence
43	6	1.5	47	1 P94146	Amino acid sequenc
44	6	1.5	64	1 Y07882	Human secreted pro
45	6	1.5	67	1 W73466	Human secreted pro

ALIGNMENTS

RESULT 1	
ID W69971	W69971 standard; Protein: 401 AA.
AC W69971:	30-NOV-1998 (first entry)
DE	Human sodium-dependent phosphate cotransporter.
KW	Sodium-dependent phosphate cotransporter; human; NAFPR; cancer; myopathy;
KW	cell signalling disorder; phosphate regulation disorder; therapy.
OS	Homo sapiens.
PN	W09837198-AI.
PD	27-AUG-1998.
PF	24-FEB-1998; 003745.
PR	24-FEB-1997; US-805118.
PA	(INCY-) INCYTE PHARM INC.
PI	Bandman O. Lal P:
DR	WPI: 98-467561/40.
DR	N-PSDB: V43711.
PT	New isolated human sodium-dependent phosphate co:transporter - used
PT	to develop products for treating e.g. cancers, osteoporosis,
PT	Alzheimer's disease, diabetes, encephalopathy, myopathy,
PT	hypocalcaemia or hypoglycaemia.
PS	Claim 1, Fig 1; 6pp; English.
CC	This sequence is the human sodium-dependent phosphate cotransporter
CC	(NAFPR) of the invention. NAFPR and agonists of it can be used to treat
CC	or prevent disorders associated with decreased phosphate levels,
CC	e.g. cancers of the kidney, disorders of decreased phosphate levels
CC	including tumoural calcinosis, osteomalacia, osteoporosis, familial
CC	hypophosphataemia, rickets, cystinuria, nephrocalcinosis,
CC	glomerulonephritis, renal calculus, Alzheimer's disease, diabetes
CC	melittus, hereditary amyloidosis, myopathies including progressive
CC	external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic epilepsy,
CC	encephalopathy, cardiomyopathy, hypokalaemia, Goodpastures syndrome, and
CC	disorders of cell signalling through CAMP, ATP, NADPH and
CC	glucose-6-phosphate. Antagonists or inhibitors of NAFPR may be
CC	administered to a subject to treat or prevent disorders associated with
CC	decreased phosphate levels, e.g. hypocalcaemia, hypocalcaemia, and
CC	abnormal phosphate regulation in neurons, gastrointestinal tract and
CC	liver. The products can also be used for detection, diagnosis and drug
CC	screening.
CC	Sequence 401 AA:
QY	1 MOYDETLIPKRVPSLCSARYGIALVLFHCNFTTIAQVNVIMTVMVNSTQSOLNDS 60
DB	1 MOYDETLIPKRVPSLCSARYGIALVLFHCNFTTIAQVNVIMTVMVNSTQSOLNDS 60
QY	61 SEVLPVPSFGSLKAPKSLPAKSSILGCGAIFMRGPPPOERSRLSIALSGMLGCFET 120
DB	61 SEVLPVPSFGSLKAPKSLPAKSSILGCGAIFMRGPPPOERSRLSIALSGMLGCFET 120
QY	121 ILGGFISLTLGMPFVFYIFGVCVCCLMFVVIYDDPVPYPMISYSEKYYITSLKQO 180
DB	121 ILGGFISLTLGMPFVFYIFGVCVCCLMFVVIYDDPVPYPMISYSEKYYITSLKQO 180

Query Match 100.0%; Score 401; DB 1; Length 401;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 121 ILGGFISEFLGMPVEFYEGVGCCVCLLMFVYIDDPVSPWISTEKEYYIISSLKQ 180
 QY 181 VGSSKQPLPIKAMLRSPIMSGICGSHOMIVSTWVYIPTIYSSVYHNINDNGLLSA 240
 Db 181 VGSSKQPLPIKAMLRSPIMSGICGSHOMIVSTWVYIPTIYSSVYHNINDNGLLSA 240
 QY 241 LPIVAVVIGWGGYIADFLTKRFRLLITVRKIATIIIGSPSALLVSLPYNSGYTTAT 300
 Db 241 LPIVAVVIGWGGYIADFLTKRFRLLITVRKIATIIIGSPSALLVSLPYNSGYTTAT 300
 QY 301 ALLTSCGSLTSCGGIYINVDIAPRYSSFLMGASRGSSIAPIVYPVSGFLSQDPE 360
 Db 301 ALLTSCGSLTSCGGIYINVDIAPRYSSFLMGASRGSSIAPIVYPVSGFLSQDPE 360
 QY 361 FGMNVFFLLFAVNLGLLFYLIIFGEADVQEMAKERRKRL 401
 Db 361 FGMNVFFLLFAVNLGLLFYLIIFGEADVQEMAKERRKRL 401

RESULT 2
 W78920
 ID W78920 standard; Protein; 401 AA.
 AC 21-DEC-1998 (first entry)
 DE Human haemochromatosis protein NRP4.
 KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
 KW diagnosis; iron metabolism; NRP3; NRP4; Roret; BTF1; BTF2; BTF3;
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KW type 1 sodium transport gene.
 OS Homo sapiens.
 PN MO9814466-41.
 PD 09-APR-1998.
 PF 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PR 01-OCT-1996; US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kronmal GS, Laufer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 DR N-PSDB: V57910.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Claim 60: Fig 5B; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the HFE gene mutation indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophilin (BT), and can be used in the production of agonists
 CC and antagonists of BT function. Also described are: (1) a Roret gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NRP3 and NRP4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence represents
 CC NRP4.
 SQ Sequence 401 AA;

Query Match 60.1%; Score 241; DB 1; Length 401;
 Best Local Similarity 100.0%; Pred. No. 9.7e-237; Indels 0; Gaps 0;
 Matches 241; Conservative 0; Mismatches 0;

QY 161 STPWISTEKEYYIISSLKQVSSKQPLPIKAMLRSPIMSGICGSHOMIVSTWVYI 220
 Db 161 STPWISTEKEYYIISSLKQVSSKQPLPIKAMLRSPIMSGICGSHOMIVSTWVYI 220

QY 221 PTYISSVYHNINDNGLLSALPEIVAVVIGWGGYIADFLTKRFRLLITVRKIATIIIGSL 280
 Db 221 PTYISSVYHNINDNGLLSALPEIVAVVIGWGGYIADFLTKRFRLLITVRKIATIIIGSL 280
 QY 281 PSSALLVSLPYNSGYTTATALLTSCGSLTSCGGIYINVDIAPRYSSFLMGASRGFS 340
 Db 281 PSSALLVSLPYNSGYTTATALLTSCGSLTSCGGIYINVDIAPRYSSFLMGASRGFS 340
 QY 341 SIAPVIVPVSGFLSQDPEFGMRNVFFLLFAVNLGLLFYLIIFGEADVQEMAKERRKRL 400
 Db 341 SIAPVIVPVSGFLSQDPEFGMRNVFFLLFAVNLGLLFYLIIFGEADVQEMAKERRKRL 400
 QY 401 L 401
 Db 401 L 401

RESULT 3
 W78919
 ID W78919 standard; Protein; 436 AA.
 AC 21-DEC-1998 (first entry)
 DE Human haemochromatosis protein NRP3.
 KW Bovine butyrophilin; BT; human hereditary haemochromatosis; HFE;
 KW diagnosis; iron metabolism; NRP3; NRP4; Roret; BTF1; BTF2; BTF3;
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KW type 1 sodium transport gene.
 OS Homo sapiens.
 PN MO9814466-41.
 PD 09-APR-1998.
 PF 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PR 01-OCT-1996; US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kronmal GS, Laufer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 DR N-PSDB: V57909.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Claim 54: Fig 5B; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the HFE gene mutation indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophilin (BT), and can be used in the production of agonists
 CC and antagonists of BT function. Also described are: (1) a Roret gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NRP3 and NRP4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence represents
 CC NRP3.
 SQ Sequence 436 AA;

Query Match 2.2%; Score 9; DB 1; Length 436;
 Best Local Similarity 100.0%; Pred. No. 0.61;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 GCVCCLLMF 152
 Db 220 GCVCCLLMF 228

RESULT 4

R56478
ID R56478 standard; peptide; 12 AA.
AC R56478:
DT 12-SEP-1994 (first entry)
DE Vitamin E transport protein fragment.
KW Vitamin E; transport protein.
OS Rattus rattus.
PN J0602529-A.
PD 01-FEB-1994.
PF 23-MAY-1991; JP-146564.
PR (EISA) EISAI CO LTD.
PA WPI: 94-071958/09.
DR Vitamin-E-specific transport protein(s) - are effective
PT transporters of vitamin-E
PS Claim 1; Page 8; 8pp; Japanese.
CC This is the partial sequence of a vitamin E transport protein. The
CC transport protein also comprises the partial sequence described in
CC R48297.
SQ Sequence 12 AA;

Query Match 1.7%; Score 7; DB 1; Length 12;
Best Local Similarity 100.0%; Pred. No. 2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 49 NSTSPOS 55
DB 3 NSTSPOS 9

RESULT 5

R15569
ID R15569 standard; protein; 20 AA.
AC R15569:
DT 02-MAR-1992 (first entry)
DE Immunopeptide #2 derived from HPV16 E7 peptide.
KW cervical cancer; cervical intraepithelial neoplasia; CIN; wart;
KM squamous cell carcinoma; ELISA; HPV 16.
OS Synthetic.
PN W09118294-A.
PD 28-NOV-1991.
PF 13-MAY-1991; SE0335.
PR 11-MAY-1990; SE-001705.
PA (MEDS-) MEDSCAND AB.
PI Dillner J, Dillner L, Cheng HM;
DR WPI: 91-369390/50.
PT Diagnosis of human papilloma virus infection and PV-carrying
PT tumours - using synthetic peptide(s) to detect virus specific
PT antigen-antibody complexes by immunoassay
PS Disclosure; Page 38; 72pp; English.
CC This is one of two peptides which have been synthesised on the
CC basis of the amino acid sequence for the E7 protein of HPV 16. The
CC selection of peptide sequences was based on the assumption that an
CC immunoreactive region might be situated in the same relative region
CC of a protein from different HPV types. The peptides were used in
CC diagnostic immunoassays to detect HPV-infection.
CC See R15523-R15601.
SQ Sequence 20 AA;

Query Match 1.7%; Score 7; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 3.3;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
DB 11 QLNDSSE 17

RESULT 6
R31212
ID R31212 standard; peptide; 29 AA.

R31212:
AC R31212:
DT 19-MAY-1993 (first entry)
DE HPV-16 E7 peptide.
DE Human papillomavirus; E7; epitope; cervical cancer; invasive;
KW antigens; diagnostic; cytotoxic; tagged.
OS Human papillomavirus type 16.
PN EP-523391-A.
PD 20-JAN-1993.
PF 19-JUN-1992; 110367.
PR 13-JUL-1991; EP-111720.
PA (BEHW) BEHRINGER AG.
PI Gissmann L, Mueller M;
DR WPI: 93-019474/03.
PT Use of HPV-16 E6 and E7-gene derived peptide(s) and their specific
PT antibodies - for treatment and diagnosis of HPV-16
PT associated invasive cervical cancer
PS Disclosure; Page 7; 14pp; English.
CC The peptide represents an epitope of the human papillomavirus type 16
CC E7 protein, from amino acids 6-35. This epitope is useful as a
CC target for diagnosis and imaging of HPV-associated cancers, such as
CC invasive cervical cancer. Antibodies raised against this epitope,
CC tagged with cytotoxic molecules, such as cholera toxin, have
CC therapeutic potential.
CC See also R31213-5.
SQ Sequence 29 AA;

Query Match 1.7%; Score 7; DB 1; Length 29;
Best Local Similarity 100.0%; Pred. No. 4.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
DB 21 QLNDSSE 27

RESULT 7

W93288
ID W93288 standard; peptide; 30 AA.
AC W93288:
DT 27-MAY-1999 (first entry)
DE Human papillomavirus peptide fragment #2.
KW Diagnosis; skin; immune reaction; onco-protein; E6; E7.
KM Human papillomavirus.
PN DE19737409-A1.
PD 04-MAR-1999.
PF 27-AUG-1997; 037409.
PR 27-AUG-1997; DE-037409.
PA (MED1-) MEDIGENE AG.
PI Hoeft R;
DR WPI: 99-168276/15.
PT Diagnosis kit for testing skin for immune reactions against
PT onco-protein E6 and E7 - comprises onco-protein E6 and E7 and/or
PT immunologically active parts of E6 and E7 derived from human
PT papilloma virus
PS Disclosure; Column 3; 4pp; German.
CC This invention describes peptides used in a diagnosis kit for testing
CC skin for immune reactions against onco-protein E6 and E7. The method
CC of the invention comprises onco-protein E6 and E7 and/or immunologically
CC active parts of E6 and E7 derived from human papilloma virus.
SQ Sequence 30 AA;

Query Match 1.7%; Score 7; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
DB 7 QLNDSSE 13

RESULT 8

R66178
ID R66178 standard; Peptide; 57 AA.
AC R66178;
DE 10-AUG-1995 (first entry)
Mouse T118 Flt3 ligand peptide fragment.
KW Flt3 ligand; tyrosine kinase receptor ligand; ss.
OS Mus musculus.
PN W09426891-A.
PD 24-NOV-1994.
PF 18-NOV-1994; U05150.
PR 19-MAY-1993; US-065231.
PR 07-JUL-1993; US-089263.
PR 16-JUL-1993; US-092543.
PR 13-AUG-1993; US-106340.
PR 24-AUG-1993; US-112391.
PR 19-NOV-1993; US-155111.
PR 03-DEC-1993; US-162413.
PA (INRM) INSR NAT SANTE & RECH MEDICALE.
PS (SCHE) SCHERING CORP.
PI Birmbaum D, Calpepper JA, Hannum CH, Lee FD;
DR WPI; 95-006787/01.
DR N-PSDB; Q79466.
PT New ligand for the Flt3 tyrosine kinase receptor - and related
nucleic acid, vectors, host cells and antibodies, useful for
treating abnormal cell physiology and proliferation, e.g. cancer,
also for diagnosis and drug screening
PS Claim 11; page 80; 90pp; English.
A cDNA library from the human stromal cell line 29SV48, in
pME18S, was screened with an 800 bp fragment derived from
mouse clone T118. This fragment encompasses the coding region
conserved between two mouse clones, T118 and T110. Approx. 20
positive clones were selected and partially sequenced. Two
clones, S86 and S109, were found to be approx. 75% homologous
to the mouse clones over the first 163 AAs. Clone S86 continued
to show homology to T110 until the stop codon, although to a
lesser degree, for an overall homology of 66%. Clones T118 and
S109 do not show homology to each other or to the other clones
after mouse residue 163 (human residue 160). An additional mouse
clone designated MB8 has a 29 AA insert at the junction between
the common and divergent portions of the mouse ligand.
SQ Sequence 57 AA;

Query Match 1.7%; Score 7; DB 1; Length 57;
Best Local Similarity 100.0%; Pred. No. 9.2;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 TATALLT 304
| | | | |
DB 16 TATALLT 22

RESULT 9
ID R72968
AC R72968 standard; Peptide; 86 AA.
AC R72968;
DT 10-JAN-1996 (first entry)
Pig kidney cell mutarotase peptide fragment IX-15.
KW Mutarotase; pig; kidney; aldose-1 epimerase; catalytic conversion;
glucose; lambda library; probe; pBluescript; E.coli.
OS Sus scrofa.
PN J06253856-A.
PD 13-SEP-1994.
PF 26-FEB-1993; 062646.
PR 26-FEB-1993; JP-062646.
PA (AMANO) AMANO PHARM KK.
DR WPI; 95-181575/24.
PT Mutarotase gene - useful for the efficient mass production of
mutarotase
PS Example 1; Page 6; 7pp; Japanese.
CC The amino acid sequences (R72965-8) are internal protease digested
fragments of the mutarotase protein from pig kidney cells (R72966).
CC Mutarotase is an aldose-1 epimerase which catalyses the conversion

CC of alpha-D-glucose to beta-D-glucose. This fragment corresponds to
CC amino acids 145-231 of the entire protein. The peptides were generated
CC by digestion with trypsin, chymotrypsin, endopeptidase and Staphylococcus
CC V8 protease. The gene was isolated from a pig kidney cell cDNA lambda
CC library using the probe Q85558. The gene was inserted into the plasmid
CC pBluescript II and transformed into E.coli JM109 for production of the
CC protein.
SQ Sequence 86 AA;

Query Match 1.7%; Score 7; DB 1; Length 86;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 VDETLIP 9
| | | | |
DB 59 VDETLIP 65

RESULT 10
ID R22767
AC R22767 standard; peptide; 98 AA.
AC R22767;
DT 21-SEP-1992 (first entry)
DE HPV E7 peptide.
KW Human; papillomavirus; immunogenic; cervical; warts; carcinoma;
KM cancer.
OS Synthetic.
OS Homo sapiens.
PN W09205248-A.
PD 02-APR-1992.
PF 26-SEP-1991; U07081.
PR 26-SEP-1990; US-588384.
PA (BRIM) BRISTOL-MYERS SQUIB.
PI Blake J, Chen L, Hellstrom I, Hellstrom K, Hu S L,
DR WPI; 92-132119/16.
PT Immunogenic peptide(s) derived from E6 or E7 region of HPV16 -
PT and recombinant cells encoding them, useful in treatment and
PT prophylaxis of cervical warts or cancer resulting from HPV
PT infection
PS Disclosure; Fig 7; 81pp; English.
CC The peptide is the sequence of the human papillomavirus HPV 16 E7
CC nucleoprotein. Peptides corresponding to regions (pref. epitopic
CC regions) of HPV 16 E7 were synthesised by standard Merrifield
CC synthesis. Examples of such peptides are E7 1-10, 29-50 or 70-81.
CC Compositions contg. these peptides, antibodies against the peptides,
CC or recombinant cells contg. the gene encoding the immunogenic
CC peptides may be utilised in methods for inhibiting and treating HPV
CC infection and tumour initiation and progression e.g. in the
CC prevention or retardation of HPV infection.
CC See also R22766.
SQ Sequence 98 AA;

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
| | | | |
DB 27 QLNDSSE 33

RESULT 11
ID R42361
AC R42361 standard; protein; 98 AA.
AC R42361;
DT 21-MAY-1994 (first entry)
Human papillomavirus 16 E7 protein and fragments.
KW Tumours; cows; horses; donkeys; regression; udder warts; HPV16.
OS Synthetic.

PM W09320844-A.
PD 28-OCT-1993.
PF 01-APR-1993; G00679.
PR 08-APR-1992; GB-007701.
PA (CANC-) CANCER RES CAMPAIGN TECHNOLOGY.
PI Campo MS;
DR WPI: 93-351368/44.
PT Use of papilloma-virus E7 protein or fragments for the therapy of
PT papilloma-virus disease - for the regression of tumours e.g.
PT removal of warts from udders or mouth of milking cows or for
PT treatment of horses or donkeys
PS Disclosure; Fig 2; 31pp; English.
CC The sequence is that of the human papillomavirus type 16 E7 protein.
CC The protein sequence was aligned with that of bovine papillomavirus
CC type 4.
CC See also R42360.
SQ Sequence 98 AA;

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
|||||||
DB 27 QLNDSSE 33

RESULT 12
W46886
ID W46886 standard; Protein; 98 AA.
AC W46886;
DT 15-JUN-1998 (first entry)
DE Amino acid sequence of the HPV-16 E7 oncoprotein.
KW E7 oncoprotein; proliferative state; HPV; kinase activity;
KW cyclin/cyclin-dependent kinase; p21cip1; interaction; inactivation;
OS Human papillomavirus.
PM US5736318-A.
PD 07-APR-1998.
PE 17-MAR-1995; 406248.
PR 17-MAR-1995; US-406248.
PA (HARD) HARVARD COLLEGE.
PI (HARD) UNIV HARVARD.
PI Jones DL, Munger K;
DR WPI: 98-239202/21.
DR N-PSDB; V16717.
PT Evaluation of proliferative state of cells transformed with human
PT papilloma virus - by determining cyclin-dependent kinase activity
PT induced by E7 onco-protein
PS Disclosure; Columns 19-20; 14pp; English.
CC The present sequence represents Human papillomavirus (HPV), strain 16,
CC E7 oncoprotein. The proliferative state of a cell transformed with
CC HPV can be evaluated in the following manner. Cyclin/cyclin-dependent
CC kinase complexes containing protein p21cip1 (W46887-88) are isolated
CC from the transformed cell, and the HPV E7 oncoprotein added to the
CC isolated protein. Cyclin/cyclin-dependent kinase complexes are isolated
CC from an untransformed cell that is substantially homogenic with the
CC transformed cell, and the HPV E7 oncoprotein added. The kinase activities
CC of the 2 samples are measured, where a proliferating transformed cell
CC has a greater kinase activity than the untransformed cell. The method is
CC used for determining the extent of interaction and/or inactivation
CC between a cyclin/cyclin-dependent kinase inhibitor and the HPV E7
CC oncoprotein and thus evaluating the proliferative state of a transformed
CC cell.
SQ Sequence 98 AA;

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62

DB 27 QLNDSSE 33
|||||||

RESULT 13
R97562
ID R97562 standard; Protein; 172 AA.
AC R97562;
DT 11-JAN-1997 (first entry)
DE Human papilloma virus E6/E7 protein variant.
KW Human papilloma virus; E6; E7; deletion mutant; HPV;
KW immune response; humoral immune response; cellular immune response;
KW vaccine.
OS Human papilloma virus.
PM W09619496-A1.
PD 27-JUN-1996.
PE 20-DEC-1995; A00868.
PR 20-DEC-1994; AU-000157.
PA (CSLC-) CSL LTD.
PI (UYOU) UNIV QUEENSLAND.
PI Cox J, Edwards SJ, Frazer I, Webb EA;
DR WPI: 96-309518/31.
DR N-PSDB; T31834.
PT Vaccine variants of human papilloma virus antigens - contain
PT variants of E6 and/or E7 protein, pref. deletion mutants, and are
PT used to treat or prevent HPV infection
PS Example 3; Page 17; 37pp; English.
CC A variant of the human papilloma virus (HPV) E6 or E7 protein which
CC elicits a humoral and/or cellular immune response against HPV can be
CC used in vaccines against HPV or to treat HPV infection. The variant
CC is preferably a deletion mutant comprising at least half, and
CC preferably two-thirds of full length E6 or E7 protein starting from
CC the N- or C-terminal, or is a full length E6 moiety fused to a full
CC length E7 moiety. The variant optionally has a linkage moiety and a
CC foreign protein or peptide which facilitates the purification of,
CC and enhances the immunogenicity of, the fusion protein. This
CC is a fusion protein of the C-terminal end of E6 and the
CC N-terminal end of E7. The protein is also a deletion mutant
CC generated from the sequence described in T31833.
SQ Sequence 172 AA;

Query Match 1.7%; Score 7; DB 1; Length 172;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
|||||||
DB 125 QLNDSSE 131

RESULT 14
W99370
ID W99370 standard; peptide; 185 AA.
AC W99370;
DT 21-MAY-1999 (first entry)
DE Papillomavirus E7/glycoprotein signal peptide fusion protein.
KW Antitumour; immunogen; intracellular localisation; cancer; dysplasia;
KW signal peptide; papillomavirus; infection; cervix uteri.
OS Chimeric - Human papillomavirus.
OS Chimeric - Rabies virus.
PM W09903885-A1.
PD 28-JAN-1999.
PE 17-JUL-1998; F01576.
PR 18-JUL-1997; FR-009152.
PA (TRGE) TRANSGENE SA.
PI Balloul JM, Bizouarne N, Kieny MP;
DR WPI: 99-132161/11.
PT Antitumour composition containing immunogenic polypeptide with
PT altered localisation - or vector expressing this polypeptide,
PT particularly for treating or preventing cervical cancer associated
PT with human papilloma virus
PS Claim 16; Page 55-56; 74pp; French.

CC The invention relates to an antitumour composition containing a
 CC therapeutic or prophylactic agent or one or more immunogenic
 CC polypeptides, where at least one polypeptide is modified so that
 CC its intracellular localisation is different from that of the native
 CC polypeptide. This sequence represents an example of a modified
 CC immunogenic protein and corresponds to the E7 protein from human
 CC papillomavirus fused to the rabies virus glycoprotein signal peptide.
 CC The compositions, vectors and particles are used to treat or prevent
 CC cancers and tumours, specifically those associated with papillomavirus
 CC infection (e.g. cancer or low grade dysplasia of the cervix uteri).
 SQ Sequence 185 AA;

Query Match 1.7%; Score 7; DB 1; Length 185;
 Best Local Similarity 100.0%; Pred. No. 29;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
 |||||

DB 46 QLNDSSE 52

RESULT 15

Y02634
 ID Y02634 standard; Protein; 220 AA.
 AC Y02634;
 DT 22-JUN-1999 (first entry)
 DE Prot.D1/3-E7-mut(C24G,E26Q)/HPV16 protein.
 KW Chimeric; E6; E7; fusion protein; protein D; vaccine; immunotherapy;
 KM tumour; lesion; benign; malignant; virus; infection.
 OS Chimeric - Human Papillomavirus.
 PN MO9910375-A2.
 PD 04-MAR-1999.
 PE 17-AUG-1998; E05285.
 PR 22-AUG-1997; GB-017953.
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 PI Bruck C, Cabezon Silva T, Delisse AER, Gerard CMG,
 PI Lombardo-Bencheikh A;
 DR WPI: 99-190587/16.
 DR N-PSDB; X29783.
 DR Human Papilloma Virus (HPV) fusion proteins - useful in vaccines for
 PT treatment or prophylaxis of HPV induced lesions
 PS Disclosure: Fig 8; 95pp; English.
 CC This sequence represents a chimeric E6 or E7 protein or E6/E7 fusion
 CC protein from Human Papillomavirus (HPV) linked to an immunological fusion
 CC partner, in this case, a fragment of the Haemophilus influenzae B
 CC protein D. The sequence also contains a histidine tag at the C-terminus
 CC of the encoded protein. The protein can be used in a vaccine, for
 CC immuno-therapeutically treating HPV induced tumour lesions (benign or
 CC malignant) and preventing HPV viral infection.
 SQ Sequence 220 AA;

Query Match 1.7%; Score 7; DB 1; Length 220;
 Best Local Similarity 100.0%; Pred. No. 34;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
 |||||

DB 140 QLNDSSE 146

Search completed: May 25, 2000, 14:54:02
 Job time: 1740 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: May 25, 2000, 14:27:18 ; Search time 40.28 Seconds

(without alignments)
143.729 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401

Sequence: 1 MOVDELIPKRVPSICSAARY.....LIFGEADVQEWAKRKRLRL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 145308 seqs, 14437401 residues

Word size : 0

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*

2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*

3: /cgn2_6/ptodata/1/1aa/6.COMB.pep:*

4: /cgn2_6/ptodata/1/1aa/PTUS.COMB.pep:*

5: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	401	100.0	401	2	US-08-805-118-1
2	138	34.4	480	2	US-08-724-394A-11
3	9	2.2	470	2	US-08-724-394A-10
4	7	1.7	20	2	US-08-934-915-47
5	7	1.7	30	1	US-08-363-586-1
6	7	1.7	57	4	PCT-US94-05150-22
7	7	1.7	98	1	US-08-406-248-6
8	7	1.7	253	2	US-08-459-818-20
9	7	1.7	253	2	US-08-889-666-20
10	7	1.7	253	2	US-08-465-078-20
11	7	1.7	253	2	US-08-725-775-20
12	7	1.7	253	2	US-08-488-062-20
13	7	1.7	263	1	US-08-117-083-9
14	7	1.7	467	1	US-08-805-118-3
15	7	1.7	480	2	US-08-724-394A-9
16	7	1.7	560	1	US-08-647-484-2
17	7	1.7	560	1	US-08-647-481-2
18	7	1.7	560	1	US-08-430-033A-2
19	7	1.7	560	1	US-08-805-118-4
20	7	1.7	560	4	PCT-US96-05792-2
21	7	1.7	671	1	US-08-737-716-13
22	6	1.5	22	1	US-08-466-265-2
23	6	1.5	22	1	US-08-755-728-15
24	6	1.5	22	2	US-08-974-655-15
25	6	1.5	40	2	US-08-353-476-92
26	6	1.5	100	2	US-08-465-380-20
27	6	1.5	100	2	US-08-480-478-48
28	6	1.5	100	2	US-08-486-397-20

29	6	1.5	100	2	US-08-486-399-20	Sequence 20, Appl
30	6	1.5	100	2	US-08-461-965-20	Sequence 20, Appl
31	6	1.5	100	2	US-08-326-110A-48	Sequence 48, Appl
32	6	1.5	100	2	US-08-634-641-20	Sequence 20, Appl
33	6	1.5	100	3	US-09-249-471-20	Sequence 20, Appl
34	6	1.5	116	1	US-08-053-131-185	Sequence 185, App
35	6	1.5	116	2	US-08-096-762-185	Sequence 185, App
36	6	1.5	118	4	PCT-US93-07213-13	Sequence 13, Appl
37	6	1.5	160	1	US-08-479-223-11	Sequence 11, Appl
38	6	1.5	160	4	PCT-US93-00643-11	Sequence 11, Appl
39	6	1.5	190	1	US-08-441-629-16	Sequence 16, Appl
40	6	1.5	190	4	PCT-US95-09172-16	Sequence 16, Appl
41	6	1.5	193	2	US-08-438-439C-21	Sequence 21, Appl
42	6	1.5	195	1	US-08-236-427-4	Sequence 4, Appl
43	6	1.5	200	4	PCT-US93-07213-7	Sequence 7, Appl
44	6	1.5	205	1	US-08-340-820-5	Sequence 5, Appl
45	6	1.5	205	1	US-08-172-328-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-805-118-1
; Sequence 1, Application US/08805118
; Patent No. 5985604
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
; TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,118
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 401 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITV02
; CLONE: 754412
; US-08-805-118-1

Query Match 100.0%; Score 401; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 1 MOVDETLPRKVPSLCSARVIALVLFHCNFTTIAQNVIMITWAVANSTSPQSOLNDS 60
Db 1 MOVDETLPRKVPSLCSARVIALVLFHCNFTTIAQNVIMITWAVANSTSPQSOLNDS 60
OY 61 SEVLPVDSFGSLAPKSPSLPAKSSILGGQPAIMWRMGPPOERSRLCSIALSGMLGCTA 120
Db 61 SEVLPVDSFGSLAPKSPSLPAKSSILGGQPAIMWRMGPPOERSRLCSIALSGMLGCTA 120
OY 121 ILIGFISLTLGMPFVYIFGVGVCVCLLMEFVYIDDPVSYPMISSEKEYIISLSKQ 180
Db 121 ILIGFISLTLGMPFVYIFGVGVCVCLLMEFVYIDDPVSYPMISSEKEYIISLSKQ 180
OY 181 VGSCKOPLPKAMLRSLPMSICLGCFSHOMLVSTWVVYIPTIISYVHNINDNGLLSA 240
Db 181 VGSCKOPLPKAMLRSLPMSICLGCFSHOMLVSTWVVYIPTIISYVHNINDNGLLSA 240
OY 241 LPTIVAVIOMVGGYLADELTLTKKFLITRKATITIGSLPSALLVSLPYLNSGYTTAT 300
Db 241 LPTIVAVIOMVGGYLADELTLTKKFLITRKATITIGSLPSALLVSLPYLNSGYTTAT 300
OY 301 ALLTSLGSLTLGOSGIYINVDIAPRYSFSLMGASRGFSSIAPIVPTVSGFLLSODPE 360
Db 301 ALLTSLGSLTLGOSGIYINVDIAPRYSFSLMGASRGFSSIAPIVPTVSGFLLSODPE 360
OY 361 FGRANVFLLFANVNLGLFYLIIFGADVQEWAKERKRLRL 401
Db 361 FGRANVFLLFANVNLGLFYLIIFGADVQEWAKERKRLRL 401

```

RESULT 2

```

US-08-724-394A-11
; Sequence 11, Application US/08724394A
; Patent No. 5872237

```

```

; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant

```

```

; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..480
; OTHER INFORMATION: /note="NP74"
US-08-724-394A-11

```

```

Query Match 34.4%; Score 138; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 5.5e-123;
Matches 138; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 264 KRLLIVRKIAITILGSLPSALLVSLPYLNSGYITATALLTSLGSLTLGOSGIYINVD 323
Db 343 KRLLIVRKIAITILGSLPSALLVSLPYLNSGYITATALLTSLGSLTLGOSGIYINVD 402
OY 324 IAPRYSFSLMGASRGFSSIAPIVPTVSGFLLSODPEGRWVFFLLFANVNLGLFYLI 383
Db 403 IAPRYSFSLMGASRGFSSIAPIVPTVSGFLLSODPEGRWVFFLLFANVNLGLFYLI 462
OY 384 FGEADVQEWAKERKRLRL 401
Db 463 FGEADVQEWAKERKRLRL 480

```

RESULT 3

```

US-08-724-394A-10
; Sequence 10, Application US/08724394A
; Patent No. 5872237

```

```

; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region

```

Mon Jun 12 23:02:41 2000

LOCATION: 1..470
OTHER INFORMATION: /note= "NPT3"
US-08-724-394A-10

Query Match 2.2%; Score 9; DB 2; Length 470;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 144 GCVCCLMF 152
DB 220 GCVCCLMF 228

RESULT 4
US-08-934-915-47
Sequence 47, Application US/08934915

PATENT No. 5932412
GENERAL INFORMATION:
APPLICANT: DILLNER, JOAKIM
APPLICANT: DILLNER, LENA
TITLE OF INVENTION: SYNTHETIC PEPTIDES OF HUMAN
TITLE OF INVENTION: PAPILOMAVIRUS 1, 5, 6, 8,
TITLE OF INVENTION: 11, 16, 18, 31, 33 AND 56,
TITLE OF INVENTION: USEFUL IN IMMUNOSSAY FOR
NUMBER OF SEQUENCES: 193
CORRESPONDENCE ADDRESS:
ADDRESSEE: MASON & ASSOCIATES, P.A.
STREET: 17757 U.S. HWY. 19 NORTH, SUITE 500
CITY: CLEARWATER
STATE: FLORIDA
COUNTRY: U.S.A.
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: Windows 3.0
SOFTWARE: Microsoft Word 6.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/934, 915
FILING DATE: 22-SEP-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/949, 836
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: LOUISE A. FOUTCH
REGISTRATION NUMBER: 37,133
REFERENCE/DOCKET NUMBER: 1946.6
TELECOMMUNICATION INFORMATION:
TELEPHONE: 813-538-3800
TELEFAX: 813-538-3820
TELEX:
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-934-915-47

Query Match 1.7%; Score 7; DB 2; Length 20;
Best Local Similarity 100.0%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSE 62
DB 11 QLNDSE 17

RESULT 5

US-08-363-586-1
Sequence 1, Application US/08363586
PATENT No. 5629161

GENERAL INFORMATION:
APPLICANT: Mueller, Martin
APPLICANT: Glassmann, Lutz
TITLE OF INVENTION: Use of HPV-16 E6 and E7-Gene Derived
TITLE OF INVENTION: Peptides for the Diagnostic Purpose
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3315

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/363,586
FILING DATE: 23-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/909,296

FILING DATE: 09-JUL-1992
APPLICATION NUMBER: EP 91111720.8
FILING DATE: 13-JUL-1991
ATTORNEY/AGENT INFORMATION:
NAME: Hadler, Linda A.
REGISTRATION NUMBER: 33,218
REFERENCE/DOCKET NUMBER: 02481-1195-00000

TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-363-586-1

Query Match 1.7%; Score 7; DB 1; Length 30;
Best Local Similarity 100.0%; Pred. No. 7.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSE 62
DB 22 QLNDSE 28

RESULT 6
PCT-US94-05150-22
Sequence 22, Application PC/TUS9405150

GENERAL INFORMATION:
APPLICANT:
APPLICANT:

TITLE OF INVENTION: Purified Mammalian Flt3 Ligands and Agonists and Antagonist
NUMBER OF SEQUENCES: 37
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Macintosh
OPERATING SYSTEM: Macintosh 6.0.5
SOFTWARE: Microsoft Word 5.1a

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/05150
FILING DATE:
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/162,413
FILING DATE: 03-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/155,111
FILING DATE: 19-NOV-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,391
FILING DATE: 24-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/106,340
FILING DATE: 13-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/092,549
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/089,263
FILING DATE: 07-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/065,231
FILING DATE: 19-MAY-1993
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 57 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US94-05150-22

Query Match 1.7%; Score 7; DB 4; Length 57;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 298 TATALLT 304
Db 16 TATALLT 22

RESULT 7
US-08-406-248-6
Sequence 6, Application US/08406248
Patent No. 5736318
GENERAL INFORMATION:
APPLICANT: Munger, Karl
APPLICANT: Jones, D. Leanne
TITLE OF INVENTION: METHOD AND KIT FOR EVALUATING
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ann-Louise Kerner, Ph.D., Lappin & Kusner
STREET: 200 State Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/406,248
FILING DATE:
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: McDaniel, Patricia A.
REGISTRATION NUMBER: 33,194
REFERENCE/DOCKET NUMBER: HAZ-011
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-330-1311
TELEFAX: 617-330-1300
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:

LENGTH: 98 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-406-248-6

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QUNDSE 62
Db 27 QUNDSE 33

RESULT 8
US-08-459-818-20
Sequence 20, Application US/08459818
Patent No. 5851795
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damle, Nitin K.
APPLICANT: Brady, William
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FastSeq 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,818
FILING DATE: 02-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436.35US02
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-818-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QUNDSE 62
Db 182 QUNDSE 188

RESULT 9
US-08-889-666-20
Sequence 20, Application US/0889666
Patent No. 5885579

GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damlé, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/889,666
FILING DATE: 08-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-350S01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-9031
TELEFAX: 310-445-1140
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-889-666-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSSE 62
|||||||
DB 182 QLNDSSE 188

RESULT 10
US-08-465-078-20
Sequence 20, Application US/08465078
Patent No. 5685796
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damlé, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/465,078
FILING DATE: 05-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-350S01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-9031
TELEFAX: 310-445-1140
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-465-078-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSSE 62
|||||||
DB 182 QLNDSSE 188

RESULT 11
US-08-725-776-20
Sequence 20, Application US/08725776
Patent No. 5968510
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Damlé, Nitin K.
APPLICANT: Brady, William
APPLICANT: Kiener, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/725,776
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-350S01

TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-725-776-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSE 62
|||||
Db 182 QLNDSE 188

RESULT 12
US-08-488-062-20
Sequence 20, Application US/08488062
Patent No. 597318
GENERAL INFORMATION:
APPLICANT: Linsley, Peter S.
APPLICANT: Ledbetter, Jeffrey A.
APPLICANT: Dangle, Ntiln K.
APPLICANT: Brady, William
APPLICANT: Kienner, Peter A.
TITLE OF INVENTION: CTLA4 Receptor and Uses Thereof
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 11150 Santa Monica Blvd., Suite 400
CITY: Los Angeles
STATE: California
COUNTRY: USA
ZIP: 90025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,062
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/375390
FILING DATE: 18-JAN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Adriano, Sarah B.
REGISTRATION NUMBER: 34,470
REFERENCE/DOCKET NUMBER: 30436-35US01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 310-445-1140
TELEFAX: 310-445-9031
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-488-062-20

Query Match 1.7%; Score 7; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSE 62
|||||
Db 182 QLNDSE 188

RESULT 13
US-08-117-083-9
Sequence 9, Application US/08117083
Patent No. 5719054
GENERAL INFORMATION:
APPLICANT: Bournell, Michael E.
APPLICANT: Inglis, Stephen C.
APPLICANT: Munro, Alan J.
TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human
TITLE OF INVENTION: Papilloma Virus Proteins
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Walter H. Dregger
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/117,083
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Dregger, Walter H.
REGISTRATION NUMBER: 24,190
REFERENCE/DOCKET NUMBER: A-58783
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-781-1989
TELEFAX: 415-398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 263 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Protein
LOCATION: 1..263
OTHER INFORMATION: /note="Xaa refers to stop codon in
US-08-117-083-9 the open reading frame."

Query Match 1.7%; Score 7; DB 1; Length 263;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSE 62
|||||
Db 188 QLNDSE 194

RESULT 14
US-08-805-118-3
Sequence 3, Application US/08805118
Patent No. 5985604
GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT

TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0221 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 467 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 450532
US-08-805-118-3

Query Match 1.7%; Score 7; DB 2; Length 467;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 LDIAPRY 328
DB 388 LDIAPRY 394

RESULT 15
US-08-724-394A-9
Sequence 9, Application US/08724394A
Patent No. 5872237
GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Krommal, Gregory S.
APPLICANT: Laner, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Therein
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Filts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 01/957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Region
LOCATION: 1..480
OTHER INFORMATION: /note="NPT1"
US-08-724-394A-9

Query Match 1.7%; Score 7; DB 2; Length 480;
Best Local Similarity 100.0%; Pred. No. 95;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 322 LDIAPRY 328
DB 401 LDIAPRY 407

Search completed: May 25, 2000, 14:54:52
Job time: 1654 sec

OM of: US-09-391-958-1 to: GenEmbl:* out_format : pfs
Date: May 25, 2000 3:26 PM
About: Results were produced by the Gencore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEM=framed-p2n.model -DEV=xlp
-Q/cgn2_1/USFTO.spool/US09391958/rnnat_24052000_213605_8639/app-query.fasta.1
-DB=GenEmbl -OFMT=fastap -SUFFIX=oligo.rge -GAPOP=4.500
-GAPEXT=0.050 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=60.000 -XGAPEXT=60.000
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=60.000 -YGAPEXT=60.000
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdt -LIST=45 -DOCLIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09391958 -NCPU=6 -ICPU=3 -NO_XLPPX -WAIT
-THREADS=1

Search information block:
Query: US-09-391-958-1
Query length: 401
Database sequences: 882769
Database length: 486395729
Search time (sec): 737.360000

WARN: XGAPOP and YGAPEXT must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming YGAPEXT=XGAPEXT=60.000

Sequence	Strd Orig	ZScore	EScore	Len	Document
gb_pat:AR036571	+	241.00	4501.17	1,2e-242	1780 AR036571 Sequence 19 from pate
gb_p3:HSU09545	+	241.00	4501.11	1.2e-242	1795 U90545 Human sodium phosphate
gb_p12:HS283953	+	98.00	1818.74	3.0e-93	570 U23953 H. sapiens mRNA: clone CD
gb_hcg2:AL138726	+	71.00	1273.36	7.1e-63	10266 AC113872 Homo sapiens chromo
gb_hcg4:AC012145	+	11.00	144.20	5.61	149155 AC012145 Homo sapiens clone
gb_hcg3:AF182108	+	10.00	124.61	6.12	167065 AF182108 Homo sapiens clone
gb_p12:CNS019DX	+	9.00	143.45	6.17	660 AL11517 Borealis cinerea strai
gb_p12:DCCNAM	+	9.00	141.51	7.82	865 AK2977 D. carota mRNA for proli
gb_p12:AF202292	+	9.00	141.57	7.86	870 AF202292 Euscomyete sp. strai
gb_p11:NTA18135	+	9.00	141.17	8.26	922 Y18135 Nicotiana tabacum mRNA
gb_p11:AB008186	+	9.00	140.98	8.47	948 AB008186 Pisum sativum mRNA for
gb_p11:PSPCA	+	9.00	140.69	8.79	990 Y16796 Pisum sativum mRNA for
gb_ba1:AB010965	+	9.00	138.01	12.39	1467 D50851 Haloragis sp. gene f
gb_ba1:AB010965	+	9.00	138.00	12.41	1470 AB010965 Haloragis sp. gene f
gb_ba1:DS3786	+	9.00	137.99	12.43	1472 D63786 Haloragis sp. gene f
gb_ba1:HRN165	+	9.00	137.98	12.45	1475 X71884 C. crispus mRNA for beta
gb_p11:CCTB1A	+	9.00	136.59	14.87	1608 X71884 C. crispus mRNA for beta
gb_p11:MSCARBOXE	+	9.00	135.88	16.28	2006 L11333 Mouse carboxylesterase
gb_pat:AR036570	+	9.00	135.05	18.11	2266 AR036570 Sequence 18 from pate
gb_p3:HSU09544	+	9.00	135.01	18.21	2281 U90544 Human sodium phosphate
gb_p11:CCTB1B	+	9.00	132.36	25.58	3367 Y17185 C. crispus gene for beta
gb_ba1:HRN165	+	9.00	124.87	66.83	10120 X72585 C. crispus ribosomal R
gb_hcg13:AC016664	+	9.00	120.06	123.95	20543 AC016664 Homo sapiens clone H
gb_h12:CEH14J21	+	9.00	119.33	135.99	22845 AF040644 Caenorhabditis eleg
gb_ba2:AF015775	+	9.00	117.46	172.24	29950 AF015775 Bacillus subtilis yd
gb_ba2:AF006665	+	9.00	116.07	206.78	36931 AF006665 Bacillus subtilis 16
gb_p11:AF001062	+	9.00	116.06	206.99	36931 AF001062 Homo sapiens genom
gb_h11:CEH14E04	+	9.00	115.97	209.42	37470 AF125448 Caenorhabditis elegans
gb_p11:AP001063	+	9.00	114.96	247.18	45311 AF001063 Homo sapiens genom
gb_hcg4:AC015270	+	9.00	113.35	292.89	35038 AC015270 Drosophila melanog
gb_hcg2:AC003656_1	+	9.00	108.64	535.91	110000 Continuation (2 of 8) of AC0
gb_hcg2:AC003656_2	+	9.00	108.29	560.75	110000 Continuation (3 of 8) of AC0
gb_p13:HS268D13	+	9.00	108.03	560.75	115863 AL023513 Human DNA sequence
gb_hcg10:AC010141	+	9.00	107.73	537.58	120333 AC010141 Homo sapiens clone
gb_hcg10:AC013376	+	9.00	107.79	537.85	124690 AC013376 Homo sapiens chromo
gb_hcg12:AC010714	+	9.00	107.76	599.94	125189 AC010714 Drosophila melanog
gb_hcg3:AC006876	+	9.00	107.70	604.81	126356 AC006876 Caenorhabditis eleg
gb_p12:HS181N1	+	9.00	107.69	605.31	126474 U282899 Human DNA sequence fr

gb_p14:AC005728	+	9.00	107.43	625.78	131389 AC005728 Homo sapiens chr
gb_hcg4:AC011093	+	9.00	107.43	626.18	131485 AC011093 Homo sapiens clo
gb_hcg2:AL138726	+	9.00	106.99	662.52	140265 AL138726 Homo sapiens chr
gb_hcg12:AC010575	+	9.00	106.60	696.11	148447 AC010575 Drosophila melan
gb_hcg4:AC012145	-	9.00	106.57	699.00	149155 AC012145 Homo sapiens clo
gb_hcg1:HSJ181N11	+	9.00	106.42	712.87	152550 U282191 Homo sapiens chrom

seq_name: gb_pat:AR036571	
seq_documentation_block:	
LOCUS	AR036571 1780 bp DNA
DEFINITION	Sequence 19 from patent US 5872237.
ACCESSION	AR036571
VERSION	AR036571.1 GI:5953239
KEYWORDS	Unknown.
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	1 (bases 1 to 1780)
AUTHORS	Feder, J. Nathan, Krommal, G. Scott, Lauer, P. M., Ruddy, D. A., Thomas, W., Tsuchinashi, Z., and Wolff, R. K.
TITLE	Megabase transcript map: novel sequences and antibodies thereto
JOURNAL	Patent: US 5872237-A 19 16-FEB-1999;
FEATURES	Location/Qualifiers
source	1..1780
BASE COUNT	458 a 436 c 402 g 484 t
ORIGIN	

alignment_scores:	
Quality:	241.00
Ratio:	1.000
Percent Similarity:	100.000
Percent Identity:	100.000

alignment_block:	
US-09-391-958-1 x AR036571	..

Align seg 1/1 to: AR036571 from: 1 to: 1780	
---	--

161	SerTyrProTrpIleSerThrSerGluValGluValTyrIleIleSerSerle	177
857	TCCATTCATGAGTACGATGACCTGAGAAAGAAATACATCATCTCTCT	906
177	UlyGlnGlnValGlySerSerGlyGlnProLeuProIleValAlaMetL	194
907	GAACAACAGAGTGGGCTCTTAAGCAGCTTCCATCAAGCATTC	956
194	euarSerLeuProIleTrpSerIleGlyGlnGlyCysPheSerHisGln	210
957	TCAATCTTACCATTTGGTCCATATGTTAGGCTGTTTCAGCATCAA	1006
211	TrpLeuValSerThrMetValTyrIleProThrTyrIleSerSerVa	227
1007	TGGTACTTACGACCAATGTTGATACATACCAATTCATCAAGCTCTG	1056
227	TyrHisValAlaSerIleArgAspAsnGlyLeuSerAlaLeuProhet	244
1057	GTACCAATGTTACATCAAGACATGACATGCTTATCGCCCTTCTTTA	1106
244	LeValAlaTrpValIleGlyMetValGlyGlyTyrLeuAlaAspPheLeu	260
1107	TGTGCTCTGGGTCATATGATGATGTTGGAAGCTATTCGACAGATTCTT	1156
261	LeuThrValSerPheArgLeuIleThrValArgGlyIleAlaThrIlele	277
1157	CTAACCAAAAAGTTTACATCATCATCATGTCAGGAAATTCACCAATTT	1206
277	uGlySerLeuProSerSerAlaLeuIleValSerLeuProTyrLeuAsn	294
1207	AGGAAGCTCTCCCTTCATCAGCATCATGCTCTTCCTTACCTCAAT	1256
294	ergGlyTyrIleThrAlaThrAlaLeuThrLeuSerGlyLeuSer	310

```

1257 CCGGCTATATACACGCAACTGCCCTGACGCTCTCTGGGATTAAGC 1306
311 ThrleucylSerGlyIleTyrlleAsnValleuaspIleAlaProar 327
1307 ACATGTGTCTACGTACGAGGATTTATATCAATGTCTAGATATGCTCCAG 1356
327 gTySerSerPheleuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
1357 GTATTCCAGTTTCTCATGGAGCATCAAGAGATTTTCGAGCATACGAC 1406
344 rovalIleValProThrValSerGlyPheleuSerGlnAspProglu 360
1407 CTGTCAATTGACCCACGTCTCAGCGATTTCTTCTAGTCAGACCTGAG 1456
361 PheGlyTrpArgAsnValPhePheleuPheAlaValAsnleuLeuG 377
1457 TTTGGGTGAGAGAAATGCTCTCTCTGCTGTTCCGTTAACCTGTTAGG 1506
377 yLeuLeuPheTyrlleuIlePheGlyGlyAlaAspValGlnGluTrpAla 394
1507 ACTACTCTTCTACCTCATATTTGGAGAGACAGATGTCAGAAATGGCTA 1556
394 ysgIuArgLysleuThrArgLeu 401
1557 AAGAGAGAAAACCTACCTCGTTTA 1579

seq_name: gb_p13:HSU90545

seq_documentation_block:
LOCUS HSU90545 1795 bp mRNA PRI 02-MAY-1997
DEFINITION Human sodium phosphate transporter (NP14) mRNA, complete cds.
ACCESSION U90545
VERSION U90545.1 GI:2062691
KEYWORDS
SOURCE
ORGANISM Homo sapiens
human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 1795)
Ruddy,D.A., Kromal,G.S., Lee,V.K., Mintier,G.A., Quintana,L.,
Domíngo,R., Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchihashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
A 1.1 megabase transcript map of the human hereditary
hemochromatosis locus
Unpublished
2 (bases 1 to 1795)
Ruddy,D.A., Kromal,G.S., Lee,V.K., Mintier,G.A., Quintana,L.,
Domíngo,R., Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchihashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
Direct Submission
Submitted (25-FEB-1997) Sequencing, Mercator Genetics, 4040
Campbell Avenue, Menlo Park, CA 94025, USA
FEATURES
source
1..1795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
377..1582
/feature="NP14"
377..1582
/feature="NP14"
377..1582
/feature="NP14"
/codon_start=1
/product="sodium phosphate transporter"
/protein_id="AAB53423.1"
/db_xref="GI:2062692"
/translation="MOVDELTPRKPSICSAARYGIALVALHCFNTTIANQYINNTM
VAMVNSTPQSDNDSSEVLPIVDSFGSLKAPKSLPAKSIIGGQFALIEKGGPOER
SKRSTALSGMLGCFETAILIGFISETLGNFVYIRFGVCCCLLMFVYIDPF
STPWISTSEKEYIISIKOVSSEKOPILKAMLSLPIWISTCIGCFBQWLVSTWV

```

```

YIPYISSYTHVNIHNDGLALPEITVAMVIGMGYLADELTKKRLITVRKIAT
IGSLPSALIVSIPYLSNGSYTATLALTSGLSGLTIOQSGIYIVLDIAPRSFTLKG
ASRGSIIAPYIVPVSFGLISQDEPFQNRVNFLLRAVNLGILFYLITREADVQEM
AKERLTRL"
BASE COUNT 473 a 436 c 402 g 484 t
ORIGIN

alignment_scores:
Quality: 241.00 Length: 241
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x HSU90545 ..
Align seg 1/1 to: HSU90545 from: 1 to: 1795

161 SerTyProThrPheSerThrSerGlyLysGlyTyrlleIleSerSerle 177
857 TCCTATTCATGATAGATAGACCTCAGAAAAAATACATCATATCTCTT 906
177 uLysGlnGlnValGlySerSerLysGlnProleuProIleLysAlaMetL 194
907 GAAACAGACAGGTGGGTCTTTAGCAGCGCTCTCCATCAAGGTATGC 956
194 euArgSerleuProIleTyPserIleCysLeuGlyCysPheSerHsgln 210
957 TCAGATCTTACCCATTGTCATATGTTTAGGCTGTTTCAGCATCA 1006
211 TrpleValSerThreValValTyrlleProThrTyrlleSerSerVa 227
1007 TGCTTAGTTCACCAATAGTGTGTATACATACCACTCATCAGCTCTGT 1056
227 lThyHsValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPhe 244
1057 GTACCATGTTATCATCAGACAGCAATGACTTATCTGCCCTCTCTT 1106
244 leValAlaTrpValIleGlyMetValGlyGlyTyrlleAlaAspPheLeu 260
1107 TTGTTGCCGTGGTCAATGAGCATGTGGGAGGCTATCTGCAGATTCT 1156
261 leuThrLysLysPheArgLeuIlePheValArgLysIleAlaThrIle 277
1157 CTACCAAAAAGTTTACCTCATCTGCTGAGGAAAATGTGCACAAATTT 1206
277 uGlySerleuProSerSerAlaLeuIleValSerleuProTyrlleuSns 294
1207 AGGAAGTCTCCCTCTTCAGCAGCATTTGTCCTGCTTACCTCAATT 1256
294 ergLysTyrlleThrAlaThrAlaLeuLeuThrleuSerCysGlyLeuSer 310
1257 CCGGCTATATACAGCAACTGCCCTGTCAGCGCTCTGCGGATTAAGC 1306
311 ThrleucylSerGlyIleTyrlleAsnValleuaspIleAlaProar 327
1307 ACATGTGTCTACGTACGAGGATTTATATCAATGTCTAGATATGCTCCAG 1356
327 gTySerSerPheleuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
1357 GTATTCCAGTTTCTCATGGAGCATCAAGAGATTTTCGAGCATACGAC 1406
344 rovalIleValProThrValSerGlyPheleuSerGlnAspProglu 360
1407 CTGTCAATTGACCCACGTCTCAGCGATTTCTTCTAGTCAGACCTGAG 1456
361 PheGlyTrpArgAsnValPhePheleuPheAlaValAsnleuLeuG 377
1457 TTTGGGTGAGAGAAATGCTCTCTCTGCTGTTCCGTTAACCTGTTAGG 1506
377 yLeuLeuPheTyrlleuIlePheGlyGlyAlaAspValGlnGluTrpAla 394
1507 ACTACTCTTCTACCTCATATTTGGAGAGACAGATGTCAGAAATGGCTA 1556

```

394 YSGIUAArglyLeuThArgleu 401
|||||
1557 AAGAGAGAAACTACTCGTTTA 1579

seq_name: gb_pr2:HS283953

seq_documentation_block:

LOCUS HS283953 570 bp mRNA PRI 16-JAN-1997

DEFINITION H.sapiens mRNA; clone CD 274.

ACCESSION Z83953

VERSION Z83953.1 GI:1783342

KEYWORDS

SOURCE

ORGANISM

human.
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 570)

AUTHORS Gasparini, P.

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 570)

AUTHORS Gasparini, P.

JOURNAL Direct Submission

Submitted (13-JAN-1997) Gasparini P., Servizio di Genetica Medica,
I.R.C.S. - Ospedale CSS Viale Cappuccini, I-71013 San Giovanni

Rotondo, Foggia, ITALY

FEATURES

source

1..570
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p22"

/clone="CD 274"

BASE COUNT 127 a 134 c 134 g 171 t 4 others

ORIGIN

alignment_scores:

Quality: 98.00 Length: 98

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x HS283953 ..

Align seg 1/1 to: HS283953 from: 1 to: 570

96 TTPGlyProProGlnGluArgSerArgLeuGlySerLeuAlaLeuSerG 112

|||||
276 TGGGGCCCTCCACAGAGACAGACACTGCGACATTCCTTATACAG 325

112 YMetLeuGlyCysPheThAlaLeuLeuGlyGlyPheLeuSerg 129

|||||
326 AATGTTACTGGATGCTTACTGCAATCCATAGTGGCTTCATTAAGTG 375

129 LurThrLeuGlyTrpProPheValPheTyrIlePheGlyGlyValGlyCys 145

|||||
376 AAACCTTGGGTGGCCCTTGTCTTATATCTTGAAGGTGGTGTGT 425

146 ValCysCysLeuLeuTrpPheValValIleTyrAspAspProValSert 162

|||||
426 GTCTGCGCTCTCTGCTGTTGTTGATTAATGATACACCCGTTTCTTA 475

162 TProTPILeSerThSerGluLysGluTyrIleIleSerSerLeuLysG 179

|||||
476 TCCATGATTAAGCACCTCAGAAAAAATACATATATCTCTTGAATC 525

179 InGlnValGlySerSerLysGlnProLeuProIleLysAlaMet 193

|||||
526 AACAGGTGGGTCTTCTTACAGACGCTCTCCATCAAGCTATG 569

seq_name: gb_htg2:AL138726

seq_documentation_block:

LOCUS AL138726 140266 bp DNA HTG 15-FEB-2000

DEFINITION Homo sapiens chromosome 6 clone RP1-139G21, *** SEQUENCING IN

PROGRESS ***, 5 unordered pieces.

ACCESSION AL138726

VERSION AL138726.2 GI:69822716

KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM

human.
Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 140266)

AUTHORS

JOURNAL

COMMENT

Submitted (15-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone
requests: clonerequest@sanger.ac.uk

On Feb 16, 2000 this sequence version replaced gi:6911804.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments. Contig_ID: 00595 length: 20679bp

Contig_ID: 00749 length: 63415bp

Contig_ID: 00902 length: 9637bp

Contig_ID: 01506 length: 27403bp

Contig_ID: 01824 length: 15932bp.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 20679: contig of 20679 bp in length

* 20680 21479: gap of 800 bp

* 21480 84894: contig of 63415 bp in length

* 84895 85694: gap of 800 bp

* 85695 95331: contig of 9637 bp in length

* 95332 96131: gap of 800 bp

* 96132 123534: contig of 27403 bp in length

* 123535 124334: gap of 800 bp

* 124335 140266: contig of 15932 bp in length.

Location/Qualifiers

1..140266

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="6"

/clone="RP1-139G21"

/clone_lib="RPC1-1"

BASE COUNT 41912 a 26597 c 25588 g 42968 t 3201 others

ORIGIN

FEATURES

source

alignment_scores:

Quality: 71.00 Length: 71

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x AL138726/rev ..

Align seg 1/1 to reverse of: AL138726 from: 1 to: 140266

12 ValProSerLeuGlySerArgTyrGlyIleAlaLeuValLeuHisPhe 28

|||||
65643 GTTCCAGTTTATGTTGCTGCTATGGAATACCCCTGCTTACATTT 65594

28 eCyAsnPherThrIleAlaGlnAsnValIleMetAsnIleThrMet 45

|||||

```

65593 CTGCAATTCACACAGATAGCACAATAATGTCATCATGACATCACCATTG 65544
45 A1AlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
|||||
65543 TAGCCATGTGTCACACAGACCAACCTCATTCCTCCAGCTCAATGATTCCTCT 65494
62 G1Ua1LeuProValAspSerPheGlyGlyLeuSerLysAlaProLysSe 78
|||||
65493 GAGGTCGCTGCTGTCATCTGTCCTTGGCTTAAAGTAAAGCCCAAGAG 65444
78 r1euproAlaLys 82
|||||
65443 TCTTCTCTGCAAG 65431
seq_name: gb_hlg4:AC012145
seq_documentation_block:
LOCUS AC012145 149155 bp DNA HTG 19-NOV-1999
DEFINITION Homo sapiens clone RP11-11K15, WORKING DRAFT SEQUENCE, 13 unordered
pieces.
ACCESSION AC012145
VERSION AC012145.2 GI:6454004
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 149155)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Homo sapiens, clone RP11-11K15
Unpublished
2 (bases 1 to 149155)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
Baldwin, J., Barina, N., Beckert, R., Boguslavsky, L., Boukhalter, B.,
Brown, A., Castle, A., Colangelo, M., Collins, S., Collymore, A.,
Cooke, P., Deatellano, K., Dewar, K., Domingo, M., Donnell, L., Doyle, K.,
Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D.,
Galagan, J., Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Locke, K., MacDonald, P., Marquis, N.,
McEwan, P., McGuck, A., McKernan, K., McLaughlin, J., Meldrum, J.,
Morrow, J., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
Peterson, K., Pollara, V., Riley, R., Roy, A., Santos, R., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
Tessier, S., Tittel, A., Vassiliev, H., Vo, A., Wheeler, J., Wu, X.,
Wyman, D., Ye, W., Zimmer, A. and Zody, M.
Direct Submission
Submitted (20-OCT-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Nov 19, 1999 this sequence version replaced gi:6087932.
Smit, A.P.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submission@genome.wi.mit.edu
----- Project Information
Center project name: L3182
Center clone name: 11_K_15
----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 108943 bases at least Q40
Consensus quality: 127281 bases at least Q30
Consensus quality: 140888 bases at least Q20
Insert size: 149155; sum-of-contrigs
Quality coverage: 4.5 in Q20 bases; sum-of-contrigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 13 contrigs. The true order of the pieces

```

```

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contrigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1927: contrig of 1927 bp in length
gap of unknown length
1928: contrig of 1382 bp in length
gap of unknown length
3310: contrig of 1462 bp in length
gap of unknown length
4772: contrig of 1179 bp in length
gap of unknown length
5951: contrig of 7018 bp in length
gap of unknown length
12969: contrig of 6874 bp in length
gap of unknown length
19843: contrig of 9438 bp in length
gap of unknown length
29280: contrig of 16233 bp in length
gap of unknown length
45514: contrig of 16167 bp in length
gap of unknown length
61681: contrig of 14859 bp in length
gap of unknown length
76540: contrig of 20209 bp in length
gap of unknown length
96749: contrig of 25513 bp in length
gap of unknown length
122262: contrig of 26894 bp in length.
Location/Qualifiers
1. 149155
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-11K15"
/clone_1lb="RPC1-11 Human Male BMC"
BASE COUNT 45294 a 30075 c 29603 g 44177 t 6 others
ORIGIN
alignment_scores:
Quality: 11.00 Length: 11
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x AC012145 ..
Align seg 1/1 to: AC012145 from: 1 to: 149155
1 MetGlnValAspGlnThrLeuLleProArgLys 11
5269 ATGCAGGTGATGAGACACTGATCCCGAGGAAA 5301
seq_name: gb_hlg4:AF182108
seq_documentation_block:
LOCUS AF182108 167065 bp DNA HTG 08-OCT-1999
DEFINITION Homo sapiens chromosome 8 clone BAC R-11n9 map 8p12.8, ***
SEQUENCING IN PROGRESS ***, in unordered pieces.
ACCESSION AF182108
VERSION AF182108.1 GI:5853325
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 167065)
Schudoy, A., Schlhabel, M., Baumgart, C., Menzel, U., Weber, J.,
Schattveyor, R. and Rosenthal, A.
Direct Submission
Submitted (31-AUG-1999) Genome Analysis, Institute of Molecular

```

COMMENT

Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
contig 1: pos. 1 - 30930 contig 2: pos. 30931 - 48337 contig 3:
pos. 48338 - 65090 contig 4: pos. 65091 - 92971 contig 5: pos.
92972 - 123996 contig 6: pos. 123997 - 134361 contig 7: pos. 134362
- 134345 contig 8: pos. 134346 - 167062 indbetween gaps of unknown
size.

* NOTE: This is a 'working draft' sequence.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

FEATURES

source
1. .167065
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="8"
/map="8p12.8"
/clone="BAC R-11N9"
514. .546
/rpt_family="MSR1"
746. .770
/note="XPOUND"
/evidence=not_experimental
882. .1153
/rpt_family="11m45"
1704. .1779
/note="MZEF"
/evidence=not_experimental
1731. .1824
/note="SST_REPEAT"
complement(2077. .2161)
/rpt_family="11m45"
2079. .2156
/rpt_family="11M45"
2080. .2210
/rpt_family="11MA5A"
complement(2091. .2154)
/rpt_family="11m42"
complement(2148. .2195)
/rpt_family="11ma9"
complement(2148. .2195)
/rpt_family="11ma5"
2779. .2806
/note="SST_REPEAT"
complement(2779. .2805)
/note="(CAAAA)n"
/rpt_family="Simple_repeat"
2904. .3033
/rpt_family="11m45"
2906. .3016
/rpt_family="11"
complement(2916. .3181)
/rpt_family="11M4C_Send"
2948. .3013
/rpt_family="11m35"
2975. .3140
/rpt_type=tandem
3041. .3133
/note="SST_REPEAT"
complement(3193. .4453)
/rpt_family="11MFC_Send"
3295. .3436
/note="SST_REPEAT"
3602. .3634
/rpt_family="11m45"
complement(3692. .3753)
/note="MZEF"
/evidence=not_experimental
3761. .3841
/rpt_family="11m45"
3941. .4184
/rpt_family="11m45"
complement(4083. .4267)
/note="GENSCAN"

exon
complement(4311. .4354)
/note="MZEF"
/evidence=not_experimental
4461. .4538
/note="XNUN_REPEAT"
4467. .4534
/note="SST_REPEAT"
4477. .4534
/note="SST_REPEAT"
4478. .4533
/rpt_type=tandem
4483. .4558
/note="SST_REPEAT"
4485. .4534
/note="SST_REPEAT"
4492. .4534
/note="SST_REPEAT"
complement(4496. .4536)
/note="(TA)n"
/rpt_family="Simple_repeat"
complement(4574. .4701)
/note="MZEF"
/evidence=not_experimental
complement(4574. .4676)
/note="GENSCAN"
5250. .5345
/rpt_family="alu"
5251. .5540
/rpt_family="ALU"
5251. .5553
/rpt_family="AluSx"
5255. .5540
/rpt_family="alu"
5258. .5380
/rpt_family="sva"
5259. .5319
/rpt_family="sva"
5372. .5451
/rpt_family="sva"
5388. .5461
/rpt_family="alu"
5405. .5464
/rpt_family="sva"
5458. .5512
/rpt_family="sva"
5529. .5562
/note="XNUN_REPEAT"
5533. .5562
/note="SST_REPEAT"
5533. .5645
/note="SST_REPEAT"
5533. .5553
/note="SST_REPEAT"
complement(6176. .6264)
/note="MZEF"
/evidence=not_experimental
complement(6176. .6336)
/note="GRAIL"
/evidence=not_experimental
complement(6425. .6488)
/note="MZEF"
/evidence=not_experimental
6482. .6548
/note="SST_REPEAT"
6624. .6686
/note="SST_REPEAT"
6624. .6658
/note="SST_REPEAT"
6624. .6698
/note="SST_REPEAT"
6630. .6698
/note="SST_REPEAT"
6639. .6675

```

AUTHORS      Genoscope.
TITLE        Direct Submission
JOURNAL      Submitted (01-SEP-1999) Genoscope - Centre National de Sequencage
              CP 5706 91057 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns
COMMENT       - web : www.genoscope.cns.fr)
              The cDNA library to be analyzed within the framework of this
              project was created using a Botrytis cinerea strain which was gro
              under conditions of nitrogen deprivation, which is the normal
              situation for B. cinerea during its development on its host plant
              The library was produced in an oriented direction, in the pBSII
              vector.
FEATURES
SOURCE       Location/Qualifiers
              1. 660
              /organism="Botryocinia fuckeliana"
              /strain="74"
              /db_xref="taxon:40559"
              /note="Genoscope sequence ID : w37C021"
BASE COUNT   157 a 152 c 148 g 203 t
ORIGIN
alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x CNS019DX
..
Align seg 1/1 to: CNS019DX from: 1 to: 660
348  prothylalserglyphenleulenser 356
|||||
442  CCCACAGCTCTCAGGCTTCCTTTGTCA 468
seq_name: gb_p12:DCPCNAM
seq_documentation_block:
LOCUS      DCPCNAM      865 bp      mRNA      PLN      30-SEP-1999
DEFINITION D carota mRNA for proliferating cell nuclear antigen (PCNA).
ACCESSION  X628277
VERSION    X62877.1 GI:18361
KEYWORDS   proliferating cell nuclear antigen.
SOURCE     carrot.
ORGANISM   Daucus carota
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
            1 (bases 1 to 865)
            Hata,S., Kouchi,H., Tanaka,Y., Minami,E., Matsumoto,T., Suzuki,I.
            and Hashimoto,J.
            Identification of carrot cDNA clones encoding a second putative
            proliferating cell-nuclear antigen, DNA polymerase delta auxiliary
            protein
            Eur. J. Biochem. 203 (3), 367-371 (1992)
JOURNAL    Eur. J. Biochem. 203 (3), 367-371 (1992)
MEDLINE    92137220
REFERENCE  2 (bases 1 to 865)
AUTHORS    Hata,S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-1991) S.Hata, Faculty of Science, Himeji
            Institute of Technology, Kamigori, Ako-gun, Hyogo 678-12, Japan
FEATURES
SOURCE     Location/Qualifiers
              1. 865
              /organism="Daucus carota"
              /cultivar="kurodagosun"
              /db_xref="taxon:4039"
BASE COUNT   245 a 168 c 222 g 230 t
ORIGIN
alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x CNS019DX
..
Align seg 1/1 to: CNS019DX from: 1 to: 660
348  prothylalserglyphenleulenser 356
|||||
442  CCCACAGCTCTCAGGCTTCCTTTGTCA 468
seq_name: gb_p12:DCPCNAM
seq_documentation_block:
LOCUS      DCPCNAM      865 bp      mRNA      PLN      30-SEP-1999
DEFINITION D carota mRNA for proliferating cell nuclear antigen (PCNA).
ACCESSION  X628277
VERSION    X62877.1 GI:18361
KEYWORDS   proliferating cell nuclear antigen.
SOURCE     carrot.
ORGANISM   Daucus carota
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
            1 (bases 1 to 865)
            Hata,S., Kouchi,H., Tanaka,Y., Minami,E., Matsumoto,T., Suzuki,I.
            and Hashimoto,J.
            Identification of carrot cDNA clones encoding a second putative
            proliferating cell-nuclear antigen, DNA polymerase delta auxiliary
            protein
            Eur. J. Biochem. 203 (3), 367-371 (1992)
JOURNAL    Eur. J. Biochem. 203 (3), 367-371 (1992)
MEDLINE    92137220
REFERENCE  2 (bases 1 to 865)
AUTHORS    Hata,S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-1991) S.Hata, Faculty of Science, Himeji
            Institute of Technology, Kamigori, Ako-gun, Hyogo 678-12, Japan
FEATURES
SOURCE     Location/Qualifiers
              1. 865
              /organism="Daucus carota"
              /cultivar="kurodagosun"
              /db_xref="taxon:4039"
BASE COUNT   245 a 168 c 222 g 230 t
ORIGIN
alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x CNS019DX
..
Align seg 1/1 to: CNS019DX from: 1 to: 660
348  prothylalserglyphenleulenser 356
|||||
442  CCCACAGCTCTCAGGCTTCCTTTGTCA 468
seq_name: gb_p12:DCPCNAM
seq_documentation_block:
LOCUS      DCPCNAM      865 bp      mRNA      PLN      30-SEP-1999
DEFINITION D carota mRNA for proliferating cell nuclear antigen (PCNA).
ACCESSION  X628277
VERSION    X62877.1 GI:18361
KEYWORDS   proliferating cell nuclear antigen.
SOURCE     carrot.
ORGANISM   Daucus carota
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
            1 (bases 1 to 865)
            Hata,S., Kouchi,H., Tanaka,Y., Minami,E., Matsumoto,T., Suzuki,I.
            and Hashimoto,J.
            Identification of carrot cDNA clones encoding a second putative
            proliferating cell-nuclear antigen, DNA polymerase delta auxiliary
            protein
            Eur. J. Biochem. 203 (3), 367-371 (1992)
JOURNAL    Eur. J. Biochem. 203 (3), 367-371 (1992)
MEDLINE    92137220
REFERENCE  2 (bases 1 to 865)
AUTHORS    Hata,S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-1991) S.Hata, Faculty of Science, Himeji
            Institute of Technology, Kamigori, Ako-gun, Hyogo 678-12, Japan
FEATURES
SOURCE     Location/Qualifiers
              1. 865
              /organism="Daucus carota"
              /cultivar="kurodagosun"
              /db_xref="taxon:4039"
BASE COUNT   245 a 168 c 222 g 230 t
ORIGIN
alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x CNS019DX
..
Align seg 1/1 to: CNS019DX from: 1 to: 660
348  prothylalserglyphenleulenser 356
|||||
442  CCCACAGCTCTCAGGCTTCCTTTGTCA 468
seq_name: gb_p12:DCPCNAM
seq_documentation_block:
LOCUS      DCPCNAM      865 bp      mRNA      PLN      30-SEP-1999
DEFINITION D carota mRNA for proliferating cell nuclear antigen (PCNA).
ACCESSION  X628277
VERSION    X62877.1 GI:18361
KEYWORDS   proliferating cell nuclear antigen.
SOURCE     carrot.
ORGANISM   Daucus carota
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
            1 (bases 1 to 865)
            Hata,S., Kouchi,H., Tanaka,Y., Minami,E., Matsumoto,T., Suzuki,I.
            and Hashimoto,J.
            Identification of carrot cDNA clones encoding a second putative
            proliferating cell-nuclear antigen, DNA polymerase delta auxiliary
            protein
            Eur. J. Biochem. 203 (3), 367-371 (1992)
JOURNAL    Eur. J. Biochem. 203 (3), 367-371 (1992)
MEDLINE    92137220
REFERENCE  2 (bases 1 to 865)
AUTHORS    Hata,S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-1991) S.Hata, Faculty of Science, Himeji
            Institute of Technology, Kamigori, Ako-gun, Hyogo 678-12, Japan
FEATURES
SOURCE     Location/Qualifiers
              1. 865
              /organism="Daucus carota"
              /cultivar="kurodagosun"
              /db_xref="taxon:4039"
BASE COUNT   245 a 168 c 222 g 230 t
ORIGIN
alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x CNS019DX
..
Align seg 1/1 to: CNS019DX from: 1 to: 660
348  prothylalserglyphenleulenser 356
|||||
442  CCCACAGCTCTCAGGCTTCCTTTGTCA 468
seq_name: gb_p12:DCPCNAM
seq_documentation_block:
LOCUS      DCPCNAM      865 bp      mRNA      PLN      30-SEP-1999
DEFINITION D carota mRNA for proliferating cell nuclear antigen (PCNA).
ACCESSION  X628277
VERSION    X62877.1 GI:18361
KEYWORDS   proliferating cell nuclear antigen.
SOURCE     carrot.
ORGANISM   Daucus carota
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
            1 (bases 1 to 865)
            Hata,S., Kouchi,H., Tanaka,Y., Minami,E., Matsumoto,T., Suzuki,I.
            and Hashimoto,J.
            Identification of carrot cDNA clones encoding a second putative
            proliferating cell-nuclear antigen, DNA polymerase delta auxiliary
            protein
            Eur. J. Biochem. 203 (3), 367-371 (1992)
JOURNAL    Eur. J. Biochem. 203 (3), 367-371 (1992)
MEDLINE    92137220
REFERENCE  2 (bases 1 to 865)
AUTHORS    Hata,S.
TITLE      Direct Submission
JOURNAL    Submitted (23-OCT-1991) S.Hata, Faculty of Science, Himeji
            Institute of Technology, Kamigori, Ako-gun, Hyogo 678-12, Japan
FEATURES
SOURCE     Location/Qualifiers
              1. 865
              /organism="Daucus carota"
              /cultivar="kurodagosun"
              /db_xref="taxon:4039"
BASE COUNT   245 a 168 c 222 g 230 t
ORIGIN
alignment_scores:
Quality:      9.00      Length:      9
Ratio:        1.000     Gaps:      0
Percent Similarity: 100.000     Percent Identity: 100.000
alignment_block:
US-09-391-958-1 x CNS019DX
..
Align seg 1/1 to: CNS019DX from: 1 to: 660
348  prothylalserglyphenleulenser 356
|||||
442  CCCACAGCTCTCAGGCTTCCTTTGTCA 468
seq_name: gb_p12:DCPCNAM
seq_documentation_block:
LOCUS      DCPCNAM      865 bp      mRNA      PLN      30-SEP-1999
DEFINITION D carota mRNA for proliferating cell nuclear antigen (PCNA).
ACCESSION  X628277
VERSION    X62877.1 GI:18361
KEYWORDS   proliferating cell nuclear antigen.
SOURCE     carrot.
ORGANISM   Daucus carota
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
            eudicots; Asteridae; euasterids II; Apiales; Apiaceae; Daucus.
            1 (bases 1 to 865)
            Hata,S., Kouchi,H., Tanaka,Y., Minami,E., Matsumoto,T., Suzuki,I.
            and Hashimoto,J.
            Identification of carrot cDNA clones encoding a second putative
            proliferating cell-nuclear antigen, DNA polymerase delta auxiliary
            protein
            Eur. J. Biochem. 203 (3), 367-371 (1992)
JOURNAL    Eur. J. Biochem. 203 (
```


Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x DCPNAM/rev ..

Align seg 1/1 to reverse of: DCPNAM from: 1 to: 865

279 SerleupProserSerAlaLeuIleVal 287
|||||
295 TCACCTGCATCATCATCAGCCTTAATAGTG 269

seq_name: gb_p12:AF202292

seq_documentation_block: 870 bp DNA PLN 27-DEC-1999

LOCUS AF202292
DEFINITION Euscomycete sp. strain cl 3-10 18S ribosomal RNA gene, partial sequence.

ACCESSION AF202292
VERSION AF202292.1 GI:6636194

KEYWORDS
SOURCE Euscomycete sp.

ORGANISM Euscomycete sp.
Eukaryota; Fungi; Ascomycota.

REFERENCE 1 (bases 1 to 870)
AUTHORS Vandenkoornhuyse, P., Martel, M. H. and Leyval, C.
TITLE Global fungal diversity associated to the roots of Arrhenatherum elatius

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 870)
AUTHORS Vandenkoornhuyse, P., Martel, M. H. and Leyval, C.
TITLE Direct Submission
JOURNAL Submitted (08-NOV-1999) CPB, CNRS UPR6831, 17, rue Notre-Dame des Pauvres, Vandoeuvre-lès-Nancy F-54500, France

FEATURES
source location/Qualifiers
1..870
/organism="Euscomycete sp."
/strain="cl 3-10"
/db_xref="taxon:110117"
<1..>870
/product="18S ribosomal RNA"

BASE COUNT 236 a 175 c 241 g 218 t
ORIGIN

RNA
/product="18S ribosomal RNA"

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AF202292/rev ..

Align seg 1/1 to reverse of: AF202292 from: 1 to: 870

6 ThirleuileProArqlylsValProser 14
|||||
743 ACCTTGATTCCTCGTAGGTCGCGAGC 717

seq_name: gb_p11:NTA18135

seq_documentation_block: 922 bp mRNA PLN 02-NOV-1998

LOCUS NTA18135
DEFINITION Nicotiana tabacum mRNA for PCNA1 protein.

ACCESSION Y18135
VERSION Y18135.1 GI:3821258
KEYWORDS PCNA1 gene.
SOURCE common tobacco.

ORGANISM Nicotiana tabacum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Asterales; Solanales; Solanaceae; Nicotiana.

REFERENCE 1 (bases 1 to 922)
AUTHORS Duong van, H., Galkwad, A. S. and Mukherjee, S.

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 922)
AUTHORS Mukherjee, S.
TITLE Direct Submission
JOURNAL Submitted (23-SEP-1998) S. Mukherjee, International Center for Genetic Engineering And Biotechnology, Plant Molecular Biology Lab, Po Box: 10504, Aruna Asaf Ali Marg, New Delhi 110067, INDIA

FEATURES
source location/Qualifiers
1..922
/organism="Nicotiana tabacum"
/db_xref="taxon:4097"
/tissue_type="leaf"
1..39
/gene="PCNA1"
1..922
/gene="PCNA1"
40..840
/gene="PCNA1"
/codon_start=1
/product="PCNA protein"
/protein_id="CAA77062.1"
/db_xref="GI:3821259"
/translation="MLEDRLVGSLLKRVLESIKELVNDANFDCSSGTGFSLOAMDSH
VALVALLRSEGEHYRCDRLNLSMGMLNNMAKMLCAGANDDITIKADGSDVTYFM
FSPYODKISDEPKMLDIDSHLGIPEVEYAIYRMPSEAFARICKDLSISGDTYVI
AVSKGVKPESTGDIQSANTVRCMTYDKPEATVIEKNEPVALQFALRYNNSFKA
TPLSSVITISLSNELPYVEYKIAEMGVRYFLAKIEDEETKPAQ"
841..922
/gene="PCNA1"

BASE COUNT 246 a 177 c 227 g 272 t
ORIGIN

3' UTR

gene

CDS

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x NTA18135/rev ..

Align seg 1/1 to reverse of: NTA18135 from: 1 to: 922

279 SerleupProserSerAlaLeuIleVal 287
|||||
329 TCACCTGCATCATCATCAGCCTTAATAGTA 303

seq_name: gb_p11:AB008186

seq_documentation_block: 948 bp mRNA PLN 17-SEP-1998

LOCUS AB008186
DEFINITION Pisum sativum mRNA for proliferating cell nuclear antigen, complete cds.

ACCESSION AB008186
VERSION AB008186.1 GI:3608174
KEYWORDS proliferating cell nuclear antigen.
SOURCE Pisum sativum cDNA to mRNA.
ORGANISM Pisum sativum

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Fabales; Fabaceae; Papilionoideae; Pisum.

REFERENCE 1 (bases 1 to 948)
AUTHORS Shimizu, S. and Mori, H.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1997) to the DDBJ/EMBL/Genbank databases. Hitoshi Mori, Nagoya University, School of Agricultural Sciences, Furo-cho Chikusa-ku, Nagoya, Aichi 464-01, Japan (E-mail: mori@bio.nagoya-u.ac.jp, Tel: 052-789-4167, Fax: 052-789-4296)

REFERENCE 2 (sites)
AUTHORS Shimizu, S. and Mori, H.
TITLE Analysis of cycles of dormancy and growth in pea axillary buds based on mRNA accumulation patterns of cell cycle-related genes

JOURNAL Plant Cell Physiol. 39 (3), 255-262 (1998)
 MEDLINE 98249622
 FEATURES Location/Qualifiers
 source 1..948
 /organism="Pisum sativum"
 /db_xref="taxon:3888"
 CDS 2..802
 /function="auxiliary protein of DNA polymerase-delta"
 /codon_start=1
 /product="proliferating cell nuclear antigen"
 /protein_id="BAA3351.1"
 /db_xref="GI:3608175"
 /translation="MLELRVQSLKLVLESIKELVNDANFDCSTGFSIQAMDSH
 VALVALLRSEGFERYRCDRLNSGMNINNAKMLKAGNDIITIKADGSDVTVM
 FESPTQDKISDEKMLMDIDSEHLGIPAEYHAIVRMPSEAFRICDLSIGDTVVI
 AVSKGEVSTKDGIGSANIVCRONTVDKREAVIEMNEPVALQFALRMNSFTKA
 TPLSSVTISLSNELPYVVEYKIAEMGVRFYLAKEIEDEETKPOA"
 BASE COUNT 251 a 179 c 225 g 293 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-391-958-1 x AB008186/rev ..

Align seg 1/1 to reverse of: AB008186 from: 1 to: 948

279 SerleupProSerSerAlaLeuIleVal 287
 |||||
 291 TCACCTTCATCATCAGCCTTAATAGTA 265

seq_name: gb_p11:PSPCNA

seq_documentation_block:

LOCUS PSPCNA 990 bp mRNA PLN 11-FEB-1999
 DEFINITION Pisum sativum mRNA for proliferating cell nuclear antigen, complete
 CDS.
 Y16796
 Y16796.1 GI:3392983
 KEYWORDS PCNA gene; proliferating cell nuclear antigen.
 SOURCE pea.
 ORGANISM Pisum sativum

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 Rosidae; Fabales; Fabaceae; Papilionoideae; Pisum.
 1 (bases 1 to 990)
 Duong van H., Gaikwad, A.S. and Mukherjee, S.K.
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 990)
 Mukherjee, S.
 TITLE Direct Submission
 JOURNAL Submitted (04-MAR-1998) S. Mukherjee, International Center for
 Genetic Engineering and Biotechnology, Plant Molecular Biology
 Lab, PO Box 10504, Aruna Asaf Ali Marg New Delhi 110067, INDIA

FEATURES
 source
 1..990
 /organism="Pisum sativum"
 /db_xref="taxon:3888"
 /clone_lib="lambda Uni-Zap"
 1..11
 /gene="PCNA"
 1..990
 /gene="PCNA"
 12..812
 /gene="PCNA"
 /codon_start=1
 /product="proliferating cell nuclear antigen"
 /protein_id="CAA76392.1"
 /db_xref="GI:4379382"

5'UTR
 gene
 CDS

/translation="MLELRVQSLKLVLESIKELVNDANFDCSTGFSIQAMDSH
 VALVALLRSEGFERYRCDRLNSGMNINNAKMLKAGNDIITIKADGSDVTVM
 FESPTQDKISDEKMLMDIDSEHLGIPAEYHAIVRMPSEAFRICDLSIGDTVVI
 AVSKGEVSTKDGIGSANIVCRONTVDKREAVIEMNEPVALQFALRMNSFTKA
 TPLSSVTISLSNELPYVVEYKIAEMGVRFYLAKEIEDEETKPOA"
 3'UTR
 BASE COUNT 283 a 181 c 229 g 297 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-391-958-1 x PSPCNA/rev ..

Align seg 1/1 to reverse of: PSPCNA from: 1 to: 990

279 SerleupProSerSerAlaLeuIleVal 287
 |||||
 301 TCACCTTCATCATCAGCCTTAATAGTA 275

seq_name: gb_ba1:D50851

seq_documentation_block:

LOCUS D50851 1467 bp DNA BCT 10-FEB-1999
 DEFINITION Haloarcula vallismortis DNA for 16S rRNA.
 ACCESSION D50851
 VERSION D50851.1 GI:1902934
 KEYWORDS 16S ribosomal RNA.
 SOURCE Haloarcula vallismortis
 ORGANISM Haloarcula vallismortis
 Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 Haloarcula.

REFERENCE 1 (bases 1 to 1467)
 Iihara, K.
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1995) to the DDBJ/EMBL/GenBank databases. Kunio
 Iihara, Nagoya University, School of Science, Department of Biology;
 Furo-cho, Chikusa-ku, Nagoya, Aichi 464-01, Japan
 (E-mail: i45031a@nuc.cc.nagoya-u.ac.jp, Tel: 052-789-2970,
 Fax: 052-789-2968)

REFERENCE 2 (bases 1 to 1467)
 Iihara, K., Watanabe, S. and Mukohata, Y.
 TITLE Haloarcula argentinensis spec. nov. and Haloarcula mukohataei sp.
 nov., two new extremely halophilic archaea collected in Argentina
 nov., two new species of extreme halophilic archaee from argentina
 3 (sites)
 JOURNAL Unpublished (1995)
 REFERENCE Iihara, K., Watanabe, S. and Tamura, T.
 TITLE Haloarcula argentinensis sp. nov. and Haloarcula mukohataei sp.
 nov., two new extremely halophilic archaea collected in Argentina
 97148980
 JOURNAL Int. J. Syst. Bacteriol. 47 (1), 73-77 (1997)
 COMMENT On Mar 21, 1997 this sequence version replaced gi:1877023.
 Sequence updated (18-Mar-1997) by: Kunio Iihara.

FEATURES
 source
 1..1467
 /organism="Haloarcula vallismortis"
 /db_xref="taxon:28442"
 1..1467
 /product="16S rRNA"
 BASE COUNT 351 a 376 c 472 g 268 t
 ORIGIN

alignment_scores:
 Quality: 9.00 Length: 9
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x D50851/rev ..
Align seg 1/1 to reverse of: D50851 from: 1 to: 1467

349 ThrValSerGlyPheLeuLeuSerGln 357
|||||
298 ACCGTCGCCGGGTTCTGCTCTACAA 272

seq_name: gb_ba1:D63786

seq_documentation_block:
LOCUS D63786 1472 bp DNA BCT 03-AUG-1999
DEFINITION Halorubrum trapanicum DNA for 16S rRNA.
ACCESSION AB010965
VERSION AB010965.1 GI:4115513
KEYWORDS 16S rRNA; 16S ribosomal RNA.
SOURCE Haloarcula sp. (strain:801030/1) DNA.
ORGANISM Haloarcula sp.
Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;
Haloarcula.
REFERENCE 1 (sites)
AUTHORS Oren, A., Ventosa, A., Gutierrez, M.C. and Kamekura, M.
TITLE Haloarcula quadrata sp. nov., a square, motile archaean isolated from a brine pool in Sinai
JOURNAL Int. J. Syst. Bacteriol. 49 Pt 3, 1149-1155 (1999)
MEDLINE 99354505
REFERENCE 2 (bases 1 to 1470)
AUTHORS Kamekura, M.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-1998) to the DDBJ/EMBL/GenBank databases.
Masahito Kamekura, Noda Institute for Scientific Research; 399
Noda, Noda, Chiba 278, Japan (E-mail:mkamekur@super.nig.ac.jp,
Tel:0471-23-5573, Fax:0471-23-5959)
Location/Qualifiers
FEATURES
source
1..1470
/organism="Haloarcula sp."
/strain="801030/1"
/db_xref="taxon:44098"
1..1470
/product="16S rRNA"
BASE COUNT 350 a 381 c 476 g 263 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AB010965/rev ..
Align seg 1/1 to reverse of: AB010965 from: 1 to: 1470

349 ThrValSerGlyPheLeuLeuSerGln 357
|||||
298 ACCGTCGCCGGGTTCTGCTCTACAA 272

seq_name: gb_ba1:D63786

seq_documentation_block:
LOCUS D63786 1472 bp DNA BCT 07-AUG-1997
DEFINITION Halorubrum trapanicum DNA for 16S rRNA.
ACCESSION D63786
VERSION D63786.1 GI:2315141
KEYWORDS 16S ribosomal RNA.
SOURCE Halorubrum trapanicum (strain:JCM8979) DNA.
ORGANISM Halorubrum trapanicum
Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;
Halorubrum.
REFERENCE 1 (bases 1 to 1472)
AUTHORS Kamekura, M.

TITLE Direct Submission
JOURNAL Submitted (06-AUG-1995) to the DDBJ/EMBL/Genbank databases.
Masahito Kamekura, Noda Institute for Scientific Research; 399
Noda, Noda, Chiba 278, Japan (E-mail:mkamekur@super.nig.ac.jp,
Tel:0471-23-5573, Fax:0471-23-5959)
REFERENCE 2 (sites)
AUTHORS Kamekura, M., Dyall-Smith, M.L., Upasani, V., Ventosa, A. and Kates, M.
TITLE Diversity of alkaliphilic halobacteria: proposals for transfer of Natronobacterium vacuolatum, Natronobacterium magadii, and Natronomonas gen. nov., respectively, as Halorubrum, Natrilaba, and Natronomonas comb. nov., Natrilaba magadii comb. nov., and Natronomonas pharaonis comb. nov., respectively
JOURNAL Int. J. Syst. Bacteriol. 47 (3), 853-857 (1997)
MEDLINE 97370612
FEATURES
Location/Qualifiers
1..1472
/organism="Halorubrum trapanicum"
/strain="JCM8979"
/db_xref="taxon:29284"
1..1472
/product="16S rRNA"
BASE COUNT 340 a 389 c 479 g 264 t
ORIGIN

alignment_scores:
Quality: 9.00 Length: 9
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x D63786/rev ..
Align seg 1/1 to reverse of: D63786 from: 1 to: 1472

349 ThrValSerGlyPheLeuLeuSerGln 357
|||||
298 ACCGTCGCCGGGTTCTGCTCTACAA 272

335 TAGCAAAAAGTCATCATGAAACATCACCACGTAGCCATGGTCAACAGC 384
 51 ThrSerProGlnSerGlnLeuAsnSperSerGlnValLeuProValAs 67
 385 ACAAGCCCTCAATCCACCTCATATATCTCTGAGGGTGGCTGTTGA 434
 67 pSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSers 84
 435 CTGATTGGTGGCTCAAGTAAAGCCCAAGAGAGTCTCTCCCAAGTCTC 484
 84 erileLeuGlyGlyGlnPheAlaIleuPgluAgtPglYProGln 100
 485 CAATACCTGGGGGTGAGTTTGCAATTGGGAAAGTGGGGCCCTCCCAA 534
 101 GluArgSerArgLeuGlySerLysAlaLeuSerGlyMetLeuLeuGly 117
 535 GAACGAGACAGACTCTGACAGATTGCTTATACAGAAATGTTACGGATG 584
 117 sPheThrAlaIleuIleuIleGlyPheIleSerGlyThrLeuGlyTrp 134
 585 CTTTACCTGCACTCATAGTGGCTTCATAGTGAACACCTGGGTGGC 634
 134 ropPheValPheTrpIlePheGlyGlyValGlyCysValGlyCysLeu 150
 635 CCTTGTCTCTATATCTTTGGAGGAGTGGCTGTCTCTCTCTCTCTC 684
 151 TrpPheValValIleTrpAspAspProValSerTrpProTrpIleSer 167
 685 TGGTTGTGTGATTTATGATGACCCCGTTCTCATCCAGATGATTAAGC 734
 167 rSerGlyLysGlyTrpIleLeuSerSerLeuLysGlnGlnValGlySers 184
 735 CTGAGAAAAAGAAATACATCATCTCTGTTGAACAAACAGGTGGTCTT 784
 184 erLysGlnProLeuProLysAlaMetLeuArgSerLeuProIleTrp 200
 785 CTAGAGAGCTCTTCCATCAAAAGCATATGCTCATCCATCTTGG 834
 201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMet 217
 835 TCCATATGTTTGGCTGTGACCCCTTCAATGATGTTAGTACACATGCT 884
 217 ValTrpIleProThrTrpIleSerSerValLysHisValAsnIleArg 234
 885 TGTATATACATACCACTTACATCATCTGTGTACCAATGTTAACACAG 934
 234 sPAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
 935 ACAATGAGACTTCTATCTGCCCTTCTTATTTGCTGGCTGATAGGC 984
 251 MetValGlyGlyTrpLeuAlaAspPheLeuThrLysLysPheArgLe 267
 985 ATGGTGGAGGCTATCTGTGACATTTCTTAAACCAAAAGTTTAACT 1034
 267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSera 284
 1035 CATCACTGAGGAGAAATTTGCCACAATTTAGAGAGTCTCCCTTTAG 1084
 284 IalLeuIleValSerLeuProTrpLeuAsnSerGlyTrpIleThrAlaThr 300
 1085 CACTATTGTGTCTCTGCCCTTACCTCAATTCGGGCTATATACAGCACT 1134
 301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuGlySglnSerGly 317
 1135 GCCTTGCTGAGGCTCTCTGGGAGTAAAGCAATTTGTCTAGGGAT 1184
 317 eTrpIleAsnValLeuAspIleAlaProArgTrpSerSerPheLeuMet 334
 1185 TTATATCAATGTCTTATATTTGCTCAAGATATTCAGTTTCTCATAG 1234
 334 LysAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
 1235 GAGCATCAAGAGGATTTTTCAGCATACACCTGATTTGATTCACCATGTC 1284

351 SerGlyPheLeuLeuSerGlnAspProGlnPheGlyTrpArgAsnValPn 367
 1285 AGCGATTTCTTCTTACTACAGACCTGAGATTTGGTGGAGGAAATGCTT 1334
 367 ePheLeuLeuPheAlaValAsnLeuGlyLeuLeuPheTrpIleLeu 384
 1335 CTTTCTGCTTTTGGCTTAACTGTAGACTACTCTTCTACTCATATN 1384
 384 heGlyGlnAlaAspValGlnGlyTrpAlaLysGlnArgLysLeuThrArg 400
 1385 TTGGAGAGACAGATGTCACAAATGGCTTAAAGAGAGAAAACTACTCGT 1434
 401 Leu 401
 1435 TTA 1437
 seq_name: N_Geneseq_36:V57910
 seq_documentation_block:
 ID V57910 standard; cDNA; 1795 BP.
 AC V57910;
 DT 21-DEC-1998 (first entry)
 DE Human haemochromatosis gene NTP4.
 KW Bovine butyrophillin; Bf; human hereditary haemochromatosis; HFE;
 diagnosis; Iron metabolism; NPT3; NPT4; Roret; BFP1; BFP2; BFP3;
 KW BFP4; BFP5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KW type 1 sodium transport gene; ss.
 OS Homo sapiens.
 FH Key
 FT CDS Location/Qualifiers
 FT 377..1582
 FT /*tag= a
 FT /product= "NTP4"
 FN WO9814466-A1.
 PD 09-APR-1998.
 PF 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PR 01-OCT-1996; US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kronmal GS, lauer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 DR P-PSDB: W78920.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Claim 58; Fig 7; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the haplotype genotype indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BFP genes, which are homologues of the milk
 CC protein butyrophillin (Bf), and can be used in the production of agonists
 CC and antagonists of Bf function. Also described are: (1) a Roret gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence encodes
 CC NTP4.
 SQ Sequence 1795 BP; 473 A; 436 C; 402 G; 484 T;

alignment_scores:
 Quality: 241.00 Length: 241
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x V57910 ..

Align seg 1/1 to: V57910 from: 1 to: 1795

```
161 SerTyrProTrpIleSerThrSerGluValGlyIleIleIleSerSerLe 177
|||||
857 TCCTATCCATGATTAAGACACCTCAGAAAGAAATACATCATCTCTCTT 906
|||||
177 ULVSGINGINValGlySerSerLysGlnProLeuProIleLysAlaMetL 194
|||||
907 GAAACACACAGCTGGGCTCTCTTAAGCAGCCTCTCCATCAAGCTATGC 956
|||||
194 eutrSerLeuProIleTrpSerIleCysLeuGlyCysPheSerHisGln 210
|||||
957 TCAGATCTCTACCATTTGGTCCATATGTTAGGCTGTTTCAGCATCAAA 1006
|||||
211 TrpLeuValSerThrMetValValTyrIleProThrTyrIleSerSerVa 227
|||||
1007 TGGTTAGTTAGCAACAATGGTTGATATACATACCAACTTACATGCTGT 1056
|||||
227 LTYRHSValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPheI 244
|||||
1057 GTACCATGTTACATCAGACACAAATGACTTCTATCTGCCCTTCTTTTA 1106
|||||
244 LeValIleATrpValIleGlyMetValGlyGlyTyrIleuAlaAspPheLeu 260
|||||
1107 TTGTGCTCTGGGTCATAGCATGGTGGAGGCTATCTGSCAGATTTCTCT 1156
|||||
261 LeuThrLysLysPheArgLeuIleThrValArgLysIleAlaThrIleLe 277
|||||
1157 CTAAACCAAAAGTTTAGACTCATCTACCTGTCAGGAAATTCGCCAATTTT 1206
|||||
277 uGlySerLeuProSerSerAlaLeuIleValSerLeuProTyrLeuAsn 294
|||||
1207 AGGAAGTCTCCCTCTTCAGACATCATGTGTCTGCTTACCTCAATY 1256
|||||
294 erGlyTyrIleThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSer 310
|||||
1257 CCGGCTATATCAGCAACAGCTGCTGACGCTCTCTCGGATTTAAGC 1306
|||||
311 ThrLeuCysGlnSerGlyIleTyrIleAsnValLeuAspIleAlaProAr 327
|||||
1307 ACAATTGCTCAGTCAGGATTTATATCAATGTCTTATATATGTTCCAAAG 1356
|||||
327 gTYrSerSerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
|||||
1357 GFATTCAGATTTCTCATGGAGCATCAAGAGATTTTCGAGCATACAC 1406
|||||
344 roValIleValProThrValSerGlyPheLeuLeuSerGlnAspProGlu 360
|||||
1407 CTGTCTATGTTACCCACATGTCAGGAGATTTCTTATGTCAGAGCCCTGAG 1456
|||||
361 PheGlyTrpArgAsnValPhePheLeuLeuPheAlaValAsnLeuLeuG 377
|||||
1457 TTTGGGTTGGAGGAATGCTCTCTCTCTGTTTCCCTTAACCTGTTAGG 1506
|||||
377 yLeuLeuPheTyrLeuIlePheGlyGlyAlaAspValGlnGluTrpAlaL 394
|||||
1507 ACTACTCTCTTACCTCATATTTTGGAGAAAGCAGATGTCCAAGAAATGGGCTA 1556
|||||
394 ySGluArgLysLeuThrArgLeu 401
|||||
1557 AAGAGAGAAAACTCATCTCGTTTA 1579
|||||
```

seq_name: N_Geneseq_36:V43712

seq_documentation_block:

ID V43712 standard; DNA; 272 BP.
AC V43712;
DT 30-NOV-1998 (first entry)
DE Sodium-dependent phosphate cotransporter coding sequence fragment.
KW Sodium-dependent phosphate cotransporter; human; NAFTR; cancer; myopathy;

KW cell signalling disorder; phosphate regulation disorder; therapy; ss.
OS Homo sapiens.
PN MO9837198-A1.
PD 27-AUG-1998.
PF 24-FEB-1998; U03745.
PR 24-FEB-1997; US-805118.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Lal P.
DR WPI: 98-467561/40.

PT New Isolated human sodium dependent phosphate co:transporter - used
PT to develop products for treating e.g. cancers, osteoporosis,
PT Alzheimer's disease, diabetes, encephalopathy, myopathy,
PT hypocalciuria or hypoglycaemia.
PS Example V: Page 47; 66pp; English.

CC This sequence represents a fragment of the DNA encoding the human
CC sodium-dependent phosphate cotransporter (NAFTR) of the invention. NAFTR
CC and agonists of it can be used to treat or prevent disorders associated
CC with decreased phosphate levels, e.g. cancers of the kidney, disorders of
CC decreased phosphate levels including tumoural calcinosis, osteomalacia,
CC osteoporosis, familial hypophosphataemia, rickets, cystinuria,
CC nephrocalcinosis, glomerulonephritis, renal calculus, Alzheimer's
CC disease, diabetes mellitus, hereditary amyloidosis, myopathies including
CC progressive external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic
CC epilepsy, encephalopathy, cardiomyopathy, hypokalaemia, Goodpastures
CC syndrome, and disorders of cell signalling through cAMP, ATP, NADPH and
CC glucose-6-phosphate. Antagonists or inhibitors of NAFTR may be
CC administered to a subject to treat or prevent disorders associated with
CC increased phosphate levels, e.g. hypocalcaemia, hypocalcaemia, and
CC abnormal phosphate regulation in neurons, gastrointestinal tract and
CC liver. The products can also be used for detection, diagnosis and drug
CC screening. 272 BP; 67 A; 56 C; 58 G; 91 T;
SQ Sequence

alignment_scores:

Quality: 85.00 Length: 85
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x V43712 ..

Align seg 1/1 to: V43712 from: 1 to: 272

```
317 IleTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeu 333
|||||
1 ATTTATATCAATGCTTAGATATGCTCAAGGATTTCCAGTTCTTCAT 50
|||||
333 tGlyAlaSerArgGlyPheSerSerIleAlaProValIleValProThrV 350
|||||
51 GGGAGCATCAAGAGATTTTCGACCATACACCTGTATGTTACCCACTG 100
|||||
350 alSerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnVal 365
|||||
101 TCAATGGATTTCTCTTAGTCAGACACCTGAGTTGGGTGGAGAAATGTC 150
|||||
367 PhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuI 383
|||||
151 TTTCTTTCGCGTTTGGCGTTAACCTGTTAGACTACTCTTCACTCAT 200
|||||
383 ePheGlyGlyAlaAspValGlnGluTrpAlaLysGlyLysLeuThrA 400
|||||
201 ATTTGGAGAAAGCAGATGTCCAAGAAATGGGCTTAAGAGAAAACTCATC 250
|||||
400 rGlu 401
|||||
251 GTTTA 255
|||||
```

seq_name: N_Geneseq_36:V57909

seq_documentation_block:

ID V57909 standard; cDNA; 2281 BP.
AC V57909;

```

PD   09-APR-1998:
PR   30-SEP-1997: U17658.
PR   07-MAY-1997: US-852495.
PA   01-OCT-1996: US-724394.
(PROC-) PROCENTIOR INC.
PI   Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI   Tsuchihashi Z, Wolff RK;
DR   WPI: 98-240014/21.
PT   Hereditary haemochromatosis gene products - used to develop products
    for the diagnosis and treatment of hereditary disorders in iron
    metabolism
PS   Example 2; Fig 8; 209pp; English.
CC   The present invention describes hereditary haemochromatosis gene
CC   products from the human haemochromatosis gene. The present sequence
CC   represents a hereditary haemochromatosis subregion from an individual
CC   unaffected by hereditary haemochromatosis (Hh). Also described is a
CC   method to determine the presence or absence of the common hereditary
CC   haemochromatosis (HFE) gene mutation in an individual comprising:
CC   (a) providing DNA or RNA from the individual; and (b) assessing the
CC   DNA or RNA for the presence or absence of a haplotype or genotype where
CC   the presence or absence of the haplotype genotype indicates the likely
CC   HFE gene sequences from the present invention can be used to develop
CC   products for use in the diagnosis and treatment of HFE. The present
CC   invention also describes Bf genes, which are homologues of the milk
CC   protein butyrophillin (Bf), and can be used in the production of agonists
CC   and antagonists of Bf function. Also described are: (1) a Roret gene
CC   which can be used to develop products for the study, diagnosis and
CC   treatment of lupus and Sjogren's syndrome; and (2) NPr3 and NPr4 genes
CC   which are homologues of a type 1 sodium transport gene, and can
CC   similarly be used for hypophosphataemia.
SQ   Sequence 235033 BP: 68786 A; 48466 C; 49441 G; 68340 T;

alignment_scores:
      Quality: 9.00          Length: 9
      Ratio: 1.000           Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x V57926 ..

Align seg 1/1 to: V57926 from: 1 to: 235033

      144 GCATGAGTAAATGACGTCGTAAGTTGGTCTTCTCCTAATGCGTTTC
      |||||
217150 GGCTGTGCTGCTGCTGCTCCTAATGCGTTTC 217176

seq_name: N_Geneseq_36:V57903

seq_documentation_block:
ID   V57903 standard; DNA; 237326 BP.
AC   V57903;
DT   21-DEC-1998 (first entry)
DE   Hereditary haemochromatosis subregion from an HH affected individual.
KW   Bovine butyrophillin; Bf; human hereditary haemochromatosis; HFE;
KW   diagnosis; Iron metabolism; NPr3; NPr4; Roret; BfF; BfF2; BfF3;
KW   BfF4; BfF5; milk protein; lupus; Sjogren's syndrome; hypophosphataemia.
OS   Homo sapiens.
SN   WO9814466-A1.
PN   WO9814466-A1.
PD   09-APR-1998.
PF   30-SEP-1997: U17658.
PR   07-MAY-1997: US-852495.
PR   01-OCT-1996: US-724394.
(PROC-) PROCENTIOR INC.
PI   Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI   Tsuchihashi Z, Wolff RK;
DR   WPI: 98-240014/21.
PT   Hereditary haemochromatosis gene products - used to develop products
    for the diagnosis and treatment of hereditary disorders in iron
    metabolism

```


Align seg 1/1 to: Q14339 from: 1 to: 1779

281 ProSerSerAlaLeuIleValSer 288
 ||||||||||||||||||
 368 CCATCCAGTGCCTCATGACTCAGT 391

seq_name: N_Geneseq_36:V82825

seq_documentation_block:

ID V82825 standard; cDNA: 2095 BP.

AC V82825;

DT 15-MAR-1999 (first entry)

DE Human GP88 autocrine growth factor cDNA.

KW GP88; granulins; epithillin; human; growth factor; autocrine; tumour;

OS cancer; viral infection; antagonist; therapy; diagnosis; ss.

PH Key Location/Qualifiers

FT CDS 13..1794

FT /tag= a

PN W09852607-A1.

PD 26-NOV-1998.

PF 22-MAY-1998; U10555.

PR 16-DEC-1997; US-991862.

PR 23-MAY-1997; US-863079.

PA (SERR/) SERRERO G.

PI Serrero G.

DR WPI; 99-045276/04.

DR P-PSDB; W85475.

PT Composition containing antagonist of growth factor GP88 - useful for

PT treating cancer and viral diseases and also for diagnosing disease

PS Disclosure: Fig 9A; 86pp; English.

CC This human cDNA sequence includes a coding region for GP88 (see

CC W85474), an 88 kDa glycoprotein autocrine growth factor and

CC epithillin/granulin precursor that is expressed in a tightly

CC regulated manner in normal cells, is overexpressed and unregulated

CC acts as a stringently required growth stimulator for the

CC tumorigenic cells. Inhibition of GP88 expression or action in the

CC properties of the overproducing cells. Antagonists to GP88 are

CC used to treat diseases associated with increased expression of

CC GP88, particularly cancer but also viral infections. Fragments of

CC GP88 are used to raise specific antibodies (used as antagonists,

CC compounds to GP88-expressing cells) and to screen for antibodies.

CC Antisense oligonucleotides can also be used as antagonists.

CC Methods are provided for diagnosing disease, or determining

CC susceptibility to disease, resulting from altered GP88 activity.

Sequence 2095 BP; 364 A; 681 C; 625 G; 425 T;

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x V82825 ..

Align seg 1/1 to: V82825 from: 1 to: 2095

281 ProSerSerAlaLeuIleValSer 288

|||||

380 CCATCCAGTGCCTCATGACTCAGT 403

seq_name: N_Geneseq_36:Q56794

seq_documentation_block:

ID Q56794 standard; DNA: 2124 BP.

AC Q56794;

DT 22-APR-1994 (first entry)

DE Granulin coding sequence.

KW Granulin; keratinocytes; wound healing; inhibition; peptide;

KW granulocytes; leucocytes; ss.

OS Homo sapiens.

PH Key Location/Qualifiers

FT misc_difference 1379..1381

FT /tag= a

FT /transl_except= ATG encodes Valine.

FT misc_difference 1640..1642

FT /tag= b

FT /transl_except= CAG encodes Glycine.

PN W09315195-A.

PD 05-AUG-1993.

PF 28-FEB-1992; CA0089.

PR 03-FEB-1992; US-829233.

PA (SOLO/) SOLOMON S.

PI Solomon S.

DR WPI; 93-320328/40.

DR P-PSDB; R48673.

PT New cysteine rich granulin peptide(s) from leucocyte(s) are

PT keratinocyte inhibitors useful topically for wound healing

PS Disclosure; Figure 4C; 53pp; English.

CC The granulin inhibits keratinocytes and is useful in formulations

CC for promoting the healing of wounds.

Sequence 2124 BP; 383 A; 685 C; 630 G; 426 T;

alignment_scores:

Quality: 8.00 Length: 8

Ratio: 1.000 Gaps: 0

Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x Q56794 ..

Align seg 1/1 to: Q56794 from: 1 to: 2124

281 ProSerSerAlaLeuIleValSer 288

|||||

393 CCATCCAGTGCCTCATGACTCAGT 416

seq_name: N_Geneseq_36:Q49052

seq_documentation_block:

ID Q49052 standard; DNA: 2198 BP.

AC Q49052;

DT 22-APR-1994 (first entry)

DE Granulin coding sequence.

KW Granulin; keratinocytes; wound healing; inhibition; peptide;

KW granulocytes; leucocytes; ss.

OS Homo sapiens.

PH Key Location/Qualifiers

FT misc_difference 1453..1455

FT /tag= a

FT /transl_except= ATG encodes Valine.

FT misc_difference 1714..1716

FT /tag= a

FT /transl_except= CAG encodes Glycine.

PN W09315195-A.

PD 05-AUG-1993.

PF 28-FEB-1992; CA0089.

PR 03-FEB-1992; US-829233.

PA (SOLO/) SOLOMON S.

PI Solomon S.

DR WPI; 93-320328/40.

DR P-PSDB; R48673.

PT New cysteine rich granulin peptide(s) from leucocyte(s) are

PT keratinocyte inhibitors useful topically for wound healing

PS Disclosure; Figure 4C; 53pp; English.

CC The granulin inhibits keratinocytes and is useful in formulations

CC for promoting the healing of wounds.

Sequence 2198 BP; 398 A; 708 C; 646 G; 446 T;

CC resistance (e.g. to carbenazim and to ansamitocin) on transformed
CC microorganisms. See Q55405 and Q55406 for mutated sequences.
SQ Sequence 3445 BP; 723 A; 1061 C; 892 G; 769 T;

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x Q48230

Align seg 1/1 to: Q48230 from: 1 to: 3445

278 GlySerLeuProSerSerAlaLeu 285
|||||
2360 GGTTCCTCCCTCCCTCGTCGACCTT 2383

seq_name: N_Geneseq_36:Q55405

seq_documentation_block:

ID Q55405 standard; DNA; 3445 BP.

AC Q55405;

DE 22-FEB-1994 (first entry)

KW A.chrysogenum beta-tubulin Ile(100) mutant coding sequence.

KW Beta tubulin; mutant; chemical resistance; selective marker;

KW cephalosporin; antibiotic production; ds.

OS Acromonium chrysogenum.

PH Key Location/Qualifiers

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

PS Claim 3 and Example 6; Fig 4-6 and Fig 7; 16pp; Japanese.
CC The wild-type coding sequence for beta-tubulin was isolated from
CC Acromonium chrysogenum ATCC 11550 and sequenced (Q48230). Primers
CC CTU-3 and CTU-6 were used to introduce mutations at codon 100 (Asn
CC to Ile) and 167 (Phe to Tyr), respectively. Expression of the
CC mutant proteins encoded by these sequences confers chemical
CC resistance (e.g. to carbenazim and to ansamitocin) on transformed
CC microorganisms. See also Q55406.

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x Q55405

Align seg 1/1 to: Q55405 from: 1 to: 3445

278 GlySerLeuProSerSerAlaLeu 285
|||||
2360 GGTTCCTCCCTCCCTCGTCGACCTT 2383

seq_name: N_Geneseq_36:Q55406

seq_documentation_block:

ID Q55406 standard; DNA; 3445 BP.

AC Q55406;

DE 22-FEB-1994 (first entry)

KW A.chrysogenum beta-tubulin Tyr(167) mutant coding sequence.

KW Beta tubulin; mutant; chemical resistance; selective marker;

KW cephalosporin; antibiotic production; ds.

OS Acromonium chrysogenum.

PH Key Location/Qualifiers

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

FT exon

FT intron

FT intron

FT exon

PA (TAKE) TAKEDA CHEM IND LTD.
 DR WPI: 93-277472/35.
 DR P-PSDB: R48201.
 PT DNA fragment contg. DNA coding mutant beta-tubulin - originates
 PT from *Acromonium chrysogenum*, used as selective marker for
 PT transformation of *A. chrysogenum*
 PS Claim 3 and Example 6: Fig 4-6 and Fig 7; 16pp; Japanese.
 CC The wild-type coding sequence for beta-tubulin was isolated from
 CC *Acromonium chrysogenum* ATCC 11550 and sequenced (Q48230). Primers
 CC CTU-3 and CTU-6 were used to introduce mutations at codon 100 (Asn
 CC to Ile) and 167 (Phe to Tyr), respectively. Expression of the
 CC mutant proteins encoded by these sequences confers chemical
 CC resistance (e.g. to carbendazim and to ansamitocin) on transformed
 CC microorganisms. See also Q55405.
 SQ Sequence 3445 BP; 724 A; 1061 C; 892 G; 768 T;

alignment_scores:
 Quality: 8.00 Length: 8
 Ratio: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-391-958-1 x Q55406 ..

Align seg 1/1 to: Q55406 from: 1 to: 3445
 278 GlySerLeuProSerSerAlaLeu 285
 ||||||||||||||||||||
 2360 GGTTCCTCCCTCCGTCGTCGCACCTT 2383

THIS PAGE BLANK (USPTO)

285 TGCTCGCTATGGAATAGCCCTCGTCTTACATTTC

34	lellaGlnsAnvAllelleMetAsnIleMrMetValAlaMetValAsnSer	50
335	TAGCAAAATATCATATCAATGAACATCAACCATGAGCATAGTCCAAAGC	384
51	ThrSerProGlnSerGluLeuAsnAspSerSerGluValLeuProValAs	67
385	ACAAGGCCCTCAATCCACAGTCATATGATTCCTCTAGAGTGCCTGTTGA	434
67	pSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSer	84
435	CTCATTTGGTGCCCTTAAGTAAAGCCCCAAGAGCTTCTCGCAAGAGCT	484
84	erIleLeuGlyGlyIlePheAlaIleTrpGluArgTrpGlyProProGln	100
485	CAATACTGGGGGTGACGTTTCATTTGGGAAAGTGGGGCCCTCCACAA	534
101	GluArgSerArgLeuGlySerIleAlaLeuSerGlyMetLeuLeuGly	117
535	GAACCAACACACACTCTGCACGATTCCTTATACAGAAATGTTACTGGAG	584
117	sPheThrAlaIleLeuIleGlyGlyPheIleSerGluThrLeuGlyTrp	134
585	CTTTACTGCCATCTTCATAGGTGGCTTCATATGAGAAACCTTGGGTGC	634
134	roPheValPheTrpIlePheGlyGlyAlaGlyCysValCysCysLeuLeu	150
635	CCTTTGCTCTTATATCTTTGGAGTGTGGTGCTGCTGCTGCTTC	684
151	TrpPheValValIleIleTrpAspAspProValSerTrpProTrpIleSerTh	167
685	TGCTTTGTGTGATTTATATGATGACCCGTTTCTATCCATGATGAAGAC	734
167	rSerGluGlyGluTrpIleIleSerSerLeuLysGlnAlaGlySerS	184
735	CTCAAAAAAGAAATACATCAATATCTCTGTAAGAACAAAGGTGCGTCT	784
184	erLysGlnProIleuProIleLysAlaMetLeuArgSerLeuProIleTrp	200
785	CTTAAGCAGCCCTTCCCATTAAGCATATGCTGATCTCTACCCCATTTGG	834
201	SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMetVa	217
835	TCCATATGTTTAGGCGTTCACGCCATCAATGTATGTTAGGCACAAATG	884
217	IValTrpIleProThrTrpIleSerSerValTrpHisValAsnIleArgA	234
885	TGATATACATCAACAATTCATCACTGCTCTGTATACCATGTTATACATCA	934
234	sPasnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly	250
935	ACAATGACCTTATATCGCCCTTCCTTATATGTTGGCCCGGTGCTATAGC	984
251	MetValGlyGlyTrpLeuAlaAspPheLeuLeuThrLysPheArgLe	267
985	ATGATGGAGGCTATCTGGCAGATTTCTTAAACCAAAAGTTAGACT	1034
267	uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSera	284
1035	CATCACTGTGAGGAAATATGCCAACAATTTTAAAGAGTCTCCCTCTTCAG	1084
284	IaLeuIleValSerLeuProTrpLeuAsnSerGlyTrpIleThrAlaThr	300
1085	CACCTACTTGTGCTGCTTACCTCAATTCGGGTATATATCACAGCACT	1134
301	AlaLeuLeuThrLeuSerGlyGlyLeuSerThrLeuGlyGlnSerGlyI	317
1135	GCTTGCTACGCTCTCTTGGGATTTAAGACACATGTGTCAGTCAGGAT	1184
317	eTrpIleAsnValLeuAspIleAlaProArgTrpSerSerPheLeuMetG	334
1185	TATATCATCATGTCTTAGATATGCTCCCAAGATATTCACAGTTTCTTCAGG	1234

```

334 1VALSERgrgllygheserSerrlIealaprovalIlevalProthVal 350
335 1VALSERgrgllygheserSerrlIealaprovalIlevalProthVal 350
1235 GAGCATTCAGAGGAGGATTTTCGACCATAGCACCTGTCATTGTGCCACATGTC 1284
351 SerGlypheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValAla 367
352 SerGlypheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValAla 367
1285 ACCGGATTTCCTTGTAGTCAGACCCCTCAGTTTGGGTGGAGAAATGCTCTT 1334
367 aPheLeuPheAlaValAlaAsnLeuLeuGlyLeuLeuPheTrpLeuIleP 384
368 aPheLeuPheAlaValAlaAsnLeuLeuGlyLeuLeuPheTrpLeuIleP 384
1335 CTCTTGCTGCTGTTGGCGTTAACCTGTTAGGACCTACTCTTCTACCTCATAT 1384
384 heGlyGlyAlaAlaAspValGlnGluTrpAlaIleGlyGluArgGlySerThrArg 400
385 TTGGAGAGAGCAGATGCTCCAGAAATGGCTTAAGAGAGAAATCTACTCGT 1434
401 Leu 401
1435 TTT
1435 TTTA 1437
seq_name: /cgn2_6/ptcdata/1/lna/5C_COMP.seq:US-08-724-394A-19
seq_documentation_block:
? Sequence 19, Application US/08724394A
? Patent No. 5872237
? GENERAL INFORMATION:
? APPLICANT: Feder, John N.
? APPLICANT: Krommal, Gregory S.
? APPLICANT: Iauer, Peter M.
? APPLICANT: Ruddy, David A.
? APPLICANT: Thomas, Winston
? APPLICANT: Tsuchihashi, Zenta
? APPLICANT: Wolff, Roger K.
? TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e
? NUMBER OF SEQUENCES: 31
? CORRESPONDENCE ADDRESSES:
? ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
? STREET: Two Embarcadero Center, 8th Floor
? CITY: San Francisco
? STATE: CA
? COUNTRY: USA
? ZIP: 94111-3834
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/724,394A
? FILING DATE: 01-OCT-1996
? CLASSIFICATION: 536
? ATTORNEY/AGENT INFORMATION:
? NAME: Fitts, Renee A., 35,136
? REGISTRATION NUMBER: 017957-000100
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-576-0300
? TELEFAX: 415-576-0300
? INFORMATION FOR SEQ ID NO: 19:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1780 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: not relevant
? TOPOLOGY: not relevant
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: misc.feature
? LOCATION: 1..1780
? OTHER INFORMATION: /note="cDNA 22E"
US-08-724-394A-19

```


alignment_scores: Quality: 241.00 Length: 241
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-08-724-394A-19 ..

Align seg 1/1 to: US-08-724-394A-19 from: 1 to: 1780

```
161 SerTyrProTrpIleSerThrSerGluLysGluTyrIleIleSerSerLe 177
|||||
857 TCCATCTCATGATGATAGCACCCTCAGAAAAAGAAATACATCATATCTCTCT 906
|||||
177 uLysGlnGlnValGlySerSerLysGlnProLeuProIleLysAlaMetL 194
|||||
907 GAAACAACAGAGTGGGCTCTTCTTAAAGCAGCCCTTCCATCAAAAGCTATGC 956
|||||
194 euArgSerLeuProIleTrpSerIleCysLeuGlyCysPheSerHisGln 210
|||||
957 TCAGATCTCTACCCATTGGCTCATATGTTTAGCTGTTTCAGCCATCAA 1006
|||||
211 TrpLeuValSerThrMetValValTyrIleProThrTyrIleSerSerVa 227
|||||
1007 TGGTTAGTTAGCAACAATGGTTGTATACATACCACTTACATCAGCTCTGT 1056
|||||
227 lTyHisValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPheI 244
|||||
1057 GTACCAATGTTAAACATCAAGACAATGAGACTTCTATGCTCCCTCTTTTA 1106
|||||
244 lValAlaTrpValIleGlyMetValGlyGlyTyrLeuAlaAspPheLeu 260
|||||
1107 TTGTGGCTGGGTATAGGATGATGGAGAGCTATCTGGCGATTTCCCTT 1156
|||||
261 LeuThrLysLysPheArgLeuIleThrValArgLysIleAlaThrIleLe 277
|||||
1157 CTAACCAAAAAAGTTTACATCATCATGCTGAGAAAAATGCCACAATTTT 1206
|||||
277 uGlySerLeuProSerSerAlaLeuIleValSerLeuProTyrLeuAsn 294
|||||
1207 AGGAAGTCTCCCTCTTACAGACATCATGCTGCTGCTTACCTCAAAAT 1256
|||||
294 ergLysTyrIleThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSer 310
|||||
1257 CCGGCTTATCAGACAGCAACTGCTGCTGAGCGCTCTCTGGGGAATTAAAC 1306
|||||
311 ThrLeuCysGlnSerGlyIleTyrIleAsnValLeuAspIleAlaProAr 327
|||||
1307 ACATTTGTGTGATGATGAGGATTTATATCAATGCTTACATATTGCTCAAG 1356
|||||
327 gTySerSerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaP 344
|||||
1357 GTATTTCAGATTTTCTCATGGGAGCATCAAGAGATTTTCGAGCATPAGC 1406
|||||
344 roValIleValProThrValSerGlyPheLeuLeuSerGlnAspProGlu 360
|||||
1407 CTGTCATATTGATCCCATCTGTCAGCGGATTTCTTCTTACGACGACCTGAG 1456
|||||
361 PheGlyTrpArgAsnValPhePheLeuLeuPheAlaValAsnLeuLeuG 377
|||||
1457 TTTGGGTGAGAGATGCTTCTTCTGCTGTTTCCGTTAACTGTGTAGG 1506
|||||
377 yLeuLeuPheTyrLeuIlePheGlyGlyAlaAspValGlnGluTrpAla 394
|||||
1507 ACTACTCTTTCACCTCATATTGGAAGAGATGTCACAAAGATGGGCTTA 1556
|||||
394 ySgiuArgLysLeuThrArgLeu 401
|||||
1557 AAGAGAGAAAACTCATCTGTTTA 1579
```

seq_name: /cgn2_6/ptodata/1/lna/5D.COMB.seq: US-08-805-118-5

seq_documentation_block:

```
? Sequence 5, Application US/08805118
? Patent No. 5985604
? GENERAL INFORMATION:
? APPLICANT: Lai, Preeti
? APPLICANT: Bandman, Olga
? TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
? TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESSER: Incyte Pharmaceuticals, Inc.
? STREET: 3174 Porter Drive
? CITY: Palo Alto
? STATE: CA
? COUNTRY: US
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/805,118
? FILING DATE: Filed Herewith
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0221 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? INFORMATION FOR SEQ ID NO: 5:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 272 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: BRAUTY02
? CLONE: 754412
? US-08-805-118-5
```

alignment_scores: Quality: 85.00 Length: 85
 Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-08-805-118-5 ..

Align seg 1/1 to: US-08-805-118-5 from: 1 to: 272

```
317 lIeTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMe 333
|||||
1 ATTATATATCAATGCTTGTGATATATGCTCCAAAGATATTCAGTTTCTCAT 50
|||||
333 tGAlaAserArgGlyPheSerSerIleAlaProValIleValProThrV 350
|||||
51 GGGAGCATCAAGAGGATTTTCAGAGCATGACACCTGCTCATTTGACCCACTG 100
|||||
101 TCAGTGGATTTCTCTTCTAGTCAGGACCTGAGATTGGGTGGAGGAGATGTC 150
|||||
350 aISerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnVal 366
|||||
367 PhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuI 383
|||||
151 TTCTTCTTGTGTTGGCCGTTAAACCTGTTAGGACTACTCTTCTTACCTCAT 200
|||||
383 ePheGlyGlnAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrA 400
```

|||||
201 ATTGAGAGAGATGTCAGATGGCTTAAGAGAAACTACTC 250
400 rgleu 401
|||||
251 GTTTA 255

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-724-394A-18

seq_documentation_block:

; Sequence 18, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Kronmal, Gregory S.

; APPLICANT: Lauer, Peter M.

; APPLICANT: Ruddy, David A.

; APPLICANT: Thomas, Winston

; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724.394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 18:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2266 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..2266

; OTHER INFORMATION: /note= "cDNA 22B"

US-08-724-394A-18

alignment_scores:

Quality: 9.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-08-724-394A-18

144 GlyCysValCysCysLeuLeuTrrpPhe 152

1076 GGCTGTCTCTCTCTCTCTATGGTTC 1102

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-724-394A-20

seq_documentation_block:

; Sequence 20, Application US/08724394A

; Patent No. 5872237

; GENERAL INFORMATION:

; APPLICANT: Feder, John N.

; APPLICANT: Kronmal, Gregory S.

; APPLICANT: Lauer, Peter M.

; APPLICANT: Ruddy, David A.

; APPLICANT: Thomas, Winston

; APPLICANT: Tsuchihashi, Zenta

; APPLICANT: Wolff, Roger K.

; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el

; NUMBER OF SEQUENCES: 31

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP

; STREET: Two Embarcadero Center, 8th Floor

; CITY: San Francisco

; STATE: CA

; COUNTRY: USA

; ZIP: 94111-3834

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/724.394A

; FILING DATE: 01-OCT-1996

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Fitts, Renee A.

; REGISTRATION NUMBER: 35,136

; REFERENCE/DOCKET NUMBER: 017957-000100

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415-576-0200

; TELEFAX: 415-576-0300

; INFORMATION FOR SEQ ID NO: 20:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 246240 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: not relevant

; TOPOLOGY: not relevant

; MOLECULE TYPE: cDNA

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: 1..246240

; OTHER INFORMATION: /note= "HLA-H.CONTIG"

US-08-724-394A-20

alignment_scores:

Quality: 9.00

Ratio: 1.000

Percent Similarity: 100.000

Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-08-724-394A-20/rev

144 GlyCysValCysCysLeuLeuTrrpPhe 152

22303 GGCTGTCTCTCTCTCTATGGTTC 22277

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-724-394A-21

seq_documentation_block:

; Sequence 21, Application US/08724394A

; Patent No. 5872237

```

: GENERAL INFORMATION:
: APPLICANT: Feder, John N.
: APPLICANT: Krommal, Gregory S.
: APPLICANT: Lauer, Peter M.
: APPLICANT: Ruddy, David A.
: APPLICANT: Thomas, Winston
: APPLICANT: Tsuchihashi, Zenta
: APPLICANT: Wolff, Roger K.
: TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
: TITLE OF INVENTION: Sequences and Antibodies Thereto
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
: STREET: Two Embarcadero Center, 8th Floor
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,394A
: FILING DATE: 01-OCT-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Fitts, Renee A.
: REGISTRATION NUMBER: 35,136
: REFERENCE/DOCKET NUMBER: 017957-000100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-576-0200
: TELEFAX: 415-576-0300
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 246240 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1..246240
: OTHER INFORMATION: /note="HLA-H.CONTIG"
: US-08-724-394A-21

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-724-394A-21/rev ..

Align seg 1/1 to reverse of: US-08-724-394A-21 from: 1 to: 246240

144 G1cysvalcyscysleuleutrphe 152
|||||
22303 GGCTGTCGTCTGTCCTATGTTTC 22277

seq_name: /cgn2.6/ptodata/1/lna/5C.COMB.seq:US-08-724-394A-22

seq_documentation_block:
: Sequence 22, Application US/08724394A
: Patent No. 5872237
: GENERAL INFORMATION:
: APPLICANT: Feder, John N.
: APPLICANT: Krommal, Gregory S.
: APPLICANT: Lauer, Peter M.
: APPLICANT: Ruddy, David A.
: APPLICANT: Thomas, Winston
```

```

: APPLICANT: Tsuchihashi, Zenta
: APPLICANT: Wolff, Roger K.
: TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
: TITLE OF INVENTION: Sequences and Antibodies Thereto
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
: STREET: Two Embarcadero Center, 8th Floor
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,394A
: FILING DATE: 01-OCT-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Fitts, Renee A.
: REGISTRATION NUMBER: 35,136
: REFERENCE/DOCKET NUMBER: 017957-000100
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-576-0200
: TELEFAX: 415-576-0300
: INFORMATION FOR SEQ ID NO: 22:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 246240 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: CDNA
: FEATURE:
: NAME/KEY: misc.feature
: LOCATION: 1..246240
: OTHER INFORMATION: /note="HLA-H.CONTIG"
: US-08-724-394A-22

alignment_scores:
  Quality: 9.00      Length: 9
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-724-394A-22/rev ..

Align seg 1/1 to reverse of: US-08-724-394A-22 from: 1 to: 246240

144 G1cysvalcyscysleuleutrphe 152
|||||
22303 GGCTGTCGTCTGTCCTATGTTTC 22277

seq_name: /cgn2.6/ptodata/1/lna/5D.COMB.seq:US-08-805-118-7

seq_documentation_block:
: Sequence 7, Application US/08805118
: Patent No. 5985604
: GENERAL INFORMATION:
: APPLICANT: Lal, Preeti
: APPLICANT: Bandman, Olga
: TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
: TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: US
```

ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filled Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0221 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: XLF
US-08-805-118-7

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-08-805-118-7 ..

Align seg 1/1 to: US-08-805-118-7 from: 1 to: 25

338 GlycerSerIleAlaProval 345
|||||
2 GGATTTTCAGCATAGCACCTGTC 25

seq_name: /cgn2_6/ptodata/1/ina/5A_COMB.seq:US-07-668-648-3

seq_documentation_block:

Sequence 3, Application US/07668648
Patent No. 5416192

GENERAL INFORMATION:

APPLICANT: Shoyab, Mohammed
APPLICANT: Plozman, Gregory D.

TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH

TITLE OF INVENTION: MODULATING PROTEINS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/668,648

FILING DATE: 19910819

CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-161-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212) 869-9741
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1779 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
TISSUE TYPE: Kidney
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1779
US-07-668-648-3

alignment_scores:
Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-07-668-648-3 ..

Align seg 1/1 to: US-07-668-648-3 from: 1 to: 1779

281 ProSerAlaIleValSer 288
|||||
368 CCATCCAGTGCCTGATGTCAGT 391

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-429-998-3

seq_documentation_block:

Sequence 3, Application US/08429998
Patent No. 5885961

GENERAL INFORMATION:

APPLICANT: Shoyab, Mohammed

APPLICANT: Plozman, Gregory D.

TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH

TITLE OF INVENTION: MODULATING PROTEINS

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/429,998

FILING DATE: 27-APR-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/668,648

FILING DATE: 13-MAR-1991

ATTORNEY/AGENT INFORMATION:

NAME: Mistock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 5624-161-999

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)790-9090

TELEFAX: (212) 869-9741

```

; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1779 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: kidney
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1779
; US-08-429-998-3

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-429-998-3  ..

Align seq 1/1  to: US-08-429-998-3  from: 1  to: 1779

281 ProserSerAlaLeuIleValSer 288
|||||
368 CCATCCAGTGCCTGATAGTCAGT 391

seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-431-333-3

seq_documentation_block:
; Sequence 3, Application US/08431333
; Patent No. 5965723
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Plozman, Gregory D.
; TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
; TITLE OF INVENTION: MODULATING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Penile & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/431,333
; FILING DATE: 27-APR-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/668,648
; FILING DATE: 13-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-161-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1779 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
```

```

; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: kidney
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1779
; US-08-431-333-3

alignment_scores:
  Quality: 8.00      Length: 8
  Ratio: 1.000      Gaps: 0
  Percent Similarity: 100.000  Percent Identity: 100.000

alignment_block:
  US-09-391-958-1 x US-08-431-333-3  ..

Align seq 1/1  to: US-08-431-333-3  from: 1  to: 1779

281 ProserSerAlaLeuIleValSer 288
|||||
368 CCATCCAGTGCCTGATAGTCAGT 391

seq_name: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:PCT-US91-02321-3

seq_documentation_block:
; Sequence 3, Application PC/US9102321
; GENERAL INFORMATION:
; APPLICANT: Shoyab, Mohammed
; APPLICANT: Plozman, Gregory D.
; TITLE OF INVENTION: EPITHELINS: NOVEL CYSTEINE-RICH GROWTH
; TITLE OF INVENTION: MODULATING PROTEINS
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Bristol-Myers Squibb Company
; STREET: 3005 First Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US91/02321
; FILING DATE: 19910403
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Poor, Brian W.
; REGISTRATION NUMBER: 32,928
; REFERENCE/DOCKET NUMBER: ON0071A-PC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 728-4800
; TELEFAX: (206) 448-4775
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1779 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: kidney
; FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..1779
; PCT-US91-02321-3
```

alignment_scores:

Quality: 8.00 Length: 8
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x PCT-US91-02321-3 ..

Align seg 1/1 to: PCT-US91-02321-3 from: 1 to: 1779

281 ProserSerAlaLeuIleValSer 288
|||||
368 CCATCCAGTGCCTGATGATGTCAGT 391

seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-805-118-6

seq_documentation_block:

; Sequence 6, Application US/08805118
; Patent No. 5985604
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
; TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,118
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0221 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: XLR
; US-08-805-118-6

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-08-805-118-6/rev ..

Align seg 1/1 to reverse of: US-08-805-118-6 from: 1 to: 25

330 SerPheLeuMetGlyAlaSer 336

|||||
23 AGTTTCATGGAGCATCA 3

seq_name: /cgn2_6/ptodata/1/lna/PCTUS_COMB.seq:PCT-US94-01149-51

seq_documentation_block:

; Sequence 51, Application PC/TUS9401149
; GENERAL INFORMATION:
; APPLICANT: Shatzman, Allan
; APPLICANT: Scott, Miller
; APPLICANT: Dillon, Susan B.
; APPLICANT: Kane, James
; TITLE OF INVENTION: Vaccinal Polypeptides
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation - Corporate
; STREET: U.S. Mailcode UW2220 - 709 Swedeland Road
; CITY: King of Prussia
; STATE: Pennsylvania
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/01149
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 149,150
; FILING DATE: 05-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 013,415
; FILING DATE: 01-FEB-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 108,914
; FILING DATE: 18-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 837,773
; FILING DATE: 18-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 751,896
; FILING DATE: 30-AUG-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 387,200
; FILING DATE: 28-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 238,801
; FILING DATE: 02-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 645,732
; FILING DATE: 30-AUG-1984
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50134 PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-270-5096
; TELEFAX: 215-270-5090
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 41 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; PCT-US94-01149-51

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x PCT-US94-01149-51/rev ..

Align seg 1/1 to reverse of: PCT-US94-01149-51 from: 1 to: 41

184 SerlysglnProleuproile 190
|||||
33 TCAAGCACCGTTACCGATT 13

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-325-243A-4

seq_documentation_block:

; Sequence 4, Application US/08325243A
; Patent No. 5541288
; GENERAL INFORMATION:
; APPLICANT: Nakano, Shigeru
; APPLICANT: Mabuchi, Toohiyuki
; APPLICANT: Tada, Miki
; APPLICANT: Taoda, Yasuo
; APPLICANT: Sugino, Dan
; APPLICANT: Kono, Yoshio
; APPLICANT: Nishimura, Kaoru
; APPLICANT: Okushima, Minoru
; TITLE OF INVENTION: No. 5541288el Peptide Having Elastase Inhibitory
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower, 233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/325,243A
; FILING DATE: 16-DEC-94
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 5-31758
; FILING DATE: 22-FEB-1993
; ATORNEY/AGENT INFORMATION:
; NAME: No. 5541288and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 19036/32311
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/474-6300
; TELEFAX: 312/474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 50 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: Other nucleic acid, Synthesized DNA
; US-08-325-243A-4

alignment_scores:

Quality: 7.00 Length: 7
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x US-08-325-243A-4 ..
Align seg 1/1 to: US-08-325-243A-4 from: 1 to: 50

306 SerCysglyLeuSerThrLeu 312
|||||
3 AGCTGTGGGCTTTCGACGCTG 23

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:50:46 ; Search time 51.62 Seconds
(without alignments)
455.454 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401

Sequence: 1 MOYDETLIPRKVPSLCSARY.....LIFGEADVQEMAKERKLTRL 401

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 168808 seqs, 58629743 residues

Word size : 0

Total number of hits satisfying chosen parameters: 168808

Minimum DB seq length: 0

Maximum DB seq length: 100000

Post-processing: Listing first 45 summaries

Database : PIR_63:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.0	333	2 G72002	aspartate dehydrog
2	8	2.0	376	2 D72647	hypothetical prote
3	8	2.0	836	2 T08424	tweety protein - f
4	7	1.7	67	2 C75397	hypothetical prote
5	7	1.7	68	2 S35174	cytochrome P450 (c
6	7	1.7	97	2 T15891	hypothetical prote
7	7	1.7	98	1 W7MLHS	E7 protein - human
8	7	1.7	128	2 E71061	hypothetical prote
9	7	1.7	142	2 T08514	outer membrane pro
10	7	1.7	151	2 H64325	hypothetical prote
11	7	1.7	165	2 A45192	ribosomal protein
12	7	1.7	172	2 S75094	hypothetical prote
13	7	1.7	212	2 S56410	hypothetical 23.4k
14	7	1.7	220	2 S43291	FLV3/FLK2 ligand (
15	7	1.7	220	2 A72021	fructose-1,6-bisph
16	7	1.7	235	2 A72359	conserved hypothet
17	7	1.7	245	2 G75097	hypothetical prote
18	7	1.7	249	2 I50464	MHC class II beta
19	7	1.7	266	1 CDWT	chlorophyll a/b-di
20	7	1.7	273	2 A1607	ubiquinol--cytochr
21	7	1.7	304	2 B60444	Na+/K+-exchanging
22	7	1.7	305	1 S52161	probable fructokin
23	7	1.7	310	2 A70421	fructose-1,6-bisph
24	7	1.7	324	2 DA2951	alkanal monooxygen
25	7	1.7	324	2 D38448	alkanal monooxygen
26	7	1.7	327	2 C35411	alkanal monooxygen
27	7	1.7	327	2 C37898	alkanal monooxygen
28	7	1.7	343	2 T33989	hypothetical prote
29	7	1.7	344	2 JC5942	chemokine receptor

30	7	1.7	379	1 S04040	alcohol dehydrogen
31	7	1.7	385	2 T31493	hypothetical prote
32	7	1.7	400	2 A39254	inositol-1,4-bisph
33	7	1.7	429	2 T09888	hypothetical prote
34	7	1.7	435	2 E70586	hypothetical prote
35	7	1.7	444	2 A75319	phosphoglucumutase
36	7	1.7	465	2 I39473	Na+-dependent phos
37	7	1.7	465	2 A56410	sodium/phosphate t
38	7	1.7	465	2 S69915	sodium phosphate t
39	7	1.7	467	2 A48916	sodium phosphate t
40	7	1.7	484	4 A32761	hypothetical Alzhe
41	7	1.7	492	2 B70617	hypothetical prote
42	7	1.7	512	1 A70201	virulence factor m
43	7	1.7	512	2 T38422	probable chromatin
44	7	1.7	524	2 A75588	probable protein k
45	7	1.7	537	2 T33955	hypothetical prote

ALIGNMENTS

RESULT 1
G72002
aspartate dehydrogenase - Chlamydia pneumoniae (strain CML029)
C:Species: Chlamydia pneumoniae
C>Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: G72002
R:Kaitman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, Nature Genet. 21, 385-389, 1999
A>Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: G72002
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <ARN>
A:Cross-references: GB:AE001685; GB:AE001363; NID:g4377378; PIDN:AAD19185.1; PID:g437
A:Experimental source: strain CML029
C:Genetics:
A:Gene: asd
C:Superfamily: yeast aspartate-semialdehyde dehydrogenase

Query Match
Best Local Similarity 2.0%; Score 8; DB 2; Length 333;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 182 GSSKQPLP 189
DB 206 GSSKQPLP 213

RESULT 2
D72647
hypothetical protein APE0610 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: D72647
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Ta
awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999
A>Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
A:Reference number: A72450; MUID:99310339
A:Accession: D72647
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-376 <KAW>
A:Cross-references: DBJ:AP000060; NID:g5104188; PIDN:BAA79580.1; PID:d1043366; PID:g
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE0610

Query Match
2.0%; Score 8; DB 2; Length 376;

Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 299 ATALLTIS 306
|||||||
Db 26 ATALLTIS 33

RESULT 3
T08424
tweety protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08424
R:Malaszka, R.; de Couet, H.G.; Miklos, G.L.
Proc. Natl. Acad. Sci. U.S.A. 95, 3731-3736, 1998
A:Title: Data transferability from model organisms to human beings: insights from the fu
A:Reference number: Z16415; MUID:98188272
A:Accession: T08424
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-836 <MAL>
A:Cross-references: EMBL:AF017777; NID:93004652; PID:93004653
A:Experimental source: strain Cantons
C:Genetics:
A:Gene: tty
A:Introns: 46/3; 208/3; 463/2; 520/1; 524/1

Query Match 2.0%; Score 8; DB 2; Length 836;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 375 LGLLFYL 382
|||||||
Db 63 LGLLFYL 70

RESULT 4
C75397
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: C75397
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250
A:Accession: C75397
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-67 <WHI>
A:Cross-references: GB:AE001987; GB:AE000513; NID:96459180; PID:AAF11003.1; PID:9645918
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1429
A:Map position: 1

Query Match 1.7%; Score 7; DB 2; Length 67;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 149 LMFVVI 155
|||||||
Db 12 LMFVVI 18

RESULT 5
S35174
cytochrome P450 (clone 9) - Madagascar periwinkle (fragment)

N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Catharantus roseus (Madagascar periwinkle)
C:Date: 13-Jan-1995 #sequence_revision 09-May-1997 #text_change 11-Jun-1999
C:Accession: S35174
R:Meijer, A.H.; Souer, E.; Verpoorte, R.; Hoge, J.H.C.
Plant Mol. Biol. 22, 379-383, 1993
A:Title: Isolation of cytochrome P-450 cDNA clones from the higher plant Catharantus
A:Reference number: S35168; MUID:93283641
A:Accession: S35174
A:Molecule type: mRNA
A:Residues: 1-68 <MEI>

A:Cross-references: EMBL:X69781; NID:939513; PID:CAA49436.1; PID:9395314
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; monooxygenase; oxidoreducta
E/2/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 1.7%; Score 7; DB 2; Length 68;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 SLKQOV 181
|||||||
Db 62 SLKQOV 68

RESULT 6
T15891
hypothetical protein D2096.5 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: T15891
R:Geisler, C.
Submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid D2096.
A:Reference number: Z18425
A:Accession: T15891
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-97 <GET>
A:Cross-references: EMBL:U40800; NID:91065946; PID:91065947; PID:AAA81489.1; CESP:ID
C:Genetics:
A:Gene: CESP:D2096.5
A:Introns: 49/3; 84/2

Query Match 1.7%; Score 7; DB 2; Length 97;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 305 LSCGLST 311
|||||||
Db 34 LSCGLST 40

RESULT 7
W7MLHS
E7 protein - human papillomavirus type 16
C:Species: human papillomavirus type 16
C:Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 20-Aug-1999
C:Accession: A03688; S12367; T10428
R:Seedorf, K.; Kramer, G.; Durst, M.; Suhai, S.; Rowekamp, W.G.
Virology 145, 181-185, 1985
A:Title: Human papillomavirus type 16 DNA sequence.
A:Reference number: A22355; MUID:85246220
A:Accession: A03688
A:Molecule type: DNA
A:Residues: 1-98 <SEE>
A:Cross-references: GB:K02718; NID:9333031; PID:AAA4640.1; PID:9333033
R:Barbosa, M.S.; Edmonds, C.; Fisher, C.; Schiller, J.T.; Lowy, D.R.; Vousden, K.H.
EMBO J. 9, 153-160, 1990
A:Title: The region of the HPV E7 oncoprotein homologous to adenovirus E1a and SV40 1
A:Reference number: S12367; MUID:90107938

A:Accession: S12367
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-98 <BAR>
R:Kennedy, I.M.; Haddow, J.K.; Clements, J.B.
J. Virol. 65, 2093-2097, 1991
A:Title: A negative element in the human poapillomavirus type 16 genome acts at the level
A:Reference number: Z17014; MUID:91162763
A:Accession: T10428
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-98 <REN>
A:Cross-references: EMBL:R02718; NID:g333031; PIDN:AAA46940.1; PID:g333033
C:Genetics:
A:Gene: E7
C:Superfamily: papillomavirus E7 protein
C:Keywords: DNA binding; early protein; transcription regulation; zinc finger
F:58-94/Region: zinc finger CCCC motif

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 56 QLNDSSE 62
|||||||
Db 27 QLNDSSE 33

RESULT 8
E71061
hypothetical protein PH187 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 05-Nov-1999
C:Accession: E71061
R:Kawarayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kusuda, N.; Oguchi
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; MUID:98344137
A:Accession: E71061
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-128 <KAN>
A:Cross-references: GB:AP000005; NID:g3326132; PIDN:AAA30287.1; PID:dl031230; PID:g32576
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by Genbank
C:Genetics:
A:Gene: PH187

Query Match 1.7%; Score 7; DB 2; Length 128;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 276 ILGSLPS 282
|||||||
Db 116 ILGSLPS 122

RESULT 9
T08514
outer membrane protein homolog - Enterobacter aerogenes plasmid R751
C:Species: Enterobacter aerogenes
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999
C:Accession: T08514
R:Thorsted, P.B.; Shah, D.S.; Macartney, D.; Kostelidou, K.; Thomas, C.M.
Plasmid 36, 95-111, 1996
A:Title: Conservation of the genetic switch between replication and transfer genes of in
A:Reference number: Z16434; MUID:97118926
A:Accession: T08514
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA

A:Residues: 1-142 <TRH>
A:Cross-references: EMBL:U67194; NID:g1572520; PID:g1572563
C:Genetics:
A:Gene: upf30.5
A:Genome: plasmid R751

Query Match 1.7%; Score 7; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 217 VVYIPRY 223
|||||||
Db 136 VVYIPRY 142

RESULT 10
H64325
hypothetical protein MJ0207 - Methanococcus jannaschii
C:Species: Methanococcus jannaschii
C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 22-Oct-1999
C:Accession: H64325
R:Bull, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blak
rson, J.D.; Sadow, P.W.; Overbeek, R.; Kirkness, E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek,
Science 273, 1058-1073, 1996
A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese
A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannasc
A:Reference number: A64300; MUID:96337999
A:Accession: H64325
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-151 <BUL>
A:Cross-references: GB:U67476; GB:L77117; NID:g1590942; PIDN:AMB98194.1; PID:g1498982
C:Genetics:
A:Map position: FOR198967-199422

Query Match 1.7%; Score 7; DB 2; Length 151;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 125 GF1SETL 131
|||||||
Db 122 GF1SETL 128

RESULT 11
A45192
ribosomal protein S5 - Chlamydia trachomatis
C:Species: Chlamydia trachomatis
C:Date: 01-Jan-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A45192; F71505
R:Faul, R.; Gray, G.J.; Koehnke, N.R.; Gu, L.J.
J. Bacteriol. 174, 1205-1212, 1992
A:Title: Cloning and sequence analysis of the Chlamydia trachomatis spc ribosomal pro
A:Reference number: A42645; MUID:92138612
A:Accession: A45192
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-165 <KAN>
A:Cross-references: GB:M80325; NID:g144617; PIDN:AAA23178.1; PID:g144627
A:Note: sequence extracted from NCBI backbone (NCBIN:79464, NCBI:P:79474)
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t
A:Reference number: A71570; MUID:99000809
A:Accession: F71505
A:Molecule type: DNA
A:Residues: 1-165 <ARN>
A:Cross-references: GB:AE001323; GB:AE001273; NID:g3328931; PIDN:AMC68113.1; PID:g332
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:

A:Gene: r55
C:Superfamily: Escherichia coli ribosomal protein S5
C:Keywords: protein biosynthesis; ribosome

Query Match 1.7%: Score 7; DB 2; Length 165;
Best Local Similarity 100.0%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 301 ALLTISC 307
Db 146 ALLTISC 152

RESULT 12

S75094
hypothetical protein slr0271 - Synechocystis sp. (strain PCC 6803)
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999
C:Accession: S75094
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
S
A:Reference number: S74322; MUID:97061201
A:Accession: S75094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-172 <KAN>
A:Cross-references: EMBL:D90910; GB:AB001339; NID:g1652956; PIDN:BAAL7956.1; PID:d101868
C:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: Synechocystis hypothetical protein slr0271

Query Match 1.7%: Score 7; DB 2; Length 172;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 299 ATALLTL 305
Db 11 ATALLTL 17

RESULT 13

S56410
hypothetical 23.4K protein (vacb-aidB intergenic region) - Escherichia coli
N:Alternate names: hypothetical protein o212
C:Species: Escherichia coli
C:Date: 28-Oct-1995 #sequence_revision 03-Nov-1995 #text_change 08-Oct-1999
C:Accession: S56410; D65229
R:Burland, V.; Plunkett III, G.; Sofia, H.J.; Daniels, D.L.; Blattner, F.R.
Nucleic Acids Res. 23, 2105-2119, 1995
A:Title: Analysis of the Escherichia coli genome VI: DNA sequence of the region from 92.
A:Reference number: S56314; MUID:95334362
A:Accession: S56410
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <BUR>
A:Cross-references: EMBL:U14003; NID:g1263172; PIDN:AAA97081.1; PID:g537026
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1994
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Co
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D65229
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <BLAT>
A:Cross-references: GB:AE000490; GB:U00096; NID:g2367356; PIDN:NAAC77142.1; PID:g1790628;
A:Experimental source: strain K-12, substrain MG1655

C:Genetics:
A:Gene: yjfm

Query Match 1.7%: Score 7; DB 2; Length 212;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 350 VSGFLLS 356
Db 131 VSGFLLS 137

RESULT 14

S43291
FLT3/FLK2 ligand (clone r118) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 20-Oct-1994 #sequence_revision 10-Nov-1995 #text_change 17-Mar-1999
C:Accession: S43291
R:Hannum, C.; Culpepper, J.; Campbell, D.; McClanahan, T.; Zurawski, S.; Bazan, J.F.; felt, A.; Menich, M.; Kellner, G.; Namiyama, R.; Rennick, D.; Roncarolo, M.G.; Zlotnik, Nature 368, 643-648, 1994
A:Title: Ligand for FLT3/FLK2 receptor tyrosine kinase regulates growth of haematopo
A:Reference number: S43290; MUID:94195428
A:Accession: S43291
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-220 <HAN>

Query Match 1.7%: Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 TATALLT 304
Db 179 TATALLT 185

RESULT 15

I58343
flt3 ligand isoform 5H - mouse
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 05-Nov-1999
C:Accession: I58343
R:Lyman, S.D.; James, L.; Escobar, S.; Downey, H.; de Vries, P.; Brasel, K.; Stocking, Oncogene 10, 149-157, 1995
A:Title: Identification of soluble and membrane-bound isoforms of the murine flt3 lig
A:Reference number: I58343; MUID:95124710
A:Accession: I58343
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-220 <RES>
A:Cross-references: GB:S76459; NID:g913479; PIDN:AAB33069.1; PID:g913480

Query Match 1.7%: Score 7; DB 2; Length 220;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 298 TATALLT 304
Db 179 TATALLT 185

Search completed: May 25, 2000, 14:57:56
Job time: 430 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:54:03 ; Search time 10.58 Seconds

(without alignments)
1154.295 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 401
Sequence: 1 MOVDEFLIPKVPKSLCSARY.....LIRGEADVQWMAKERKILTRL 401

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 83857 seqs, 30454973 residues

Word size : 0

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	241	60.1	401	NP74_HUMAN	O00476 homo sapien
2	9	2.2	436	NP73_HUMAN	O00624 homo sapien
3	9	1.7	98	VE7_HPV16	P03129 human papil
4	7	1.7	151	Y207_METTA	O57660 methanococc
5	7	1.7	165	R55_CHUTR	P28543 chlamydia t
6	7	1.7	172	Y271_SYNY3	P73892 synechocyst
7	7	1.7	187	BACA_AZOB3	P39438 azospirillum
8	7	1.7	209	ATPO_DROME	O24439 drosophila
9	7	1.7	212	YJFM_ECOLI	P39295 escherichia
10	7	1.7	266	CB21_WHEAT	P04784 triticum ae
11	7	1.7	273	UCRI_MAIZE	P43727 zea mays (m
12	7	1.7	280	GCVK_HSV62	P52446 herpes slmp
13	7	1.7	305	SCRK_ECOLI	P40713 escherichia
14	7	1.7	324	LXB2_PHOUJ	P23147 photorhabdu
15	7	1.7	327	LXBL_PHOUJ	P18840 photorhabdu
16	7	1.7	379	ADH3_HORVU	P10848 hordeum vul
17	7	1.7	399	INBP_HUMAN	P49441 homo sapien
18	7	1.7	400	INBP_BOVIN	P21327 bos taurus
19	7	1.7	435	YN66_MYCTU	O05833 mycobacteri
20	7	1.7	465	NP71_HUMAN	O14916 homo sapien
21	7	1.7	465	NP71_MOUSE	O61983 mus musculu
22	7	1.7	465	NP71_RABIT	O28722 oryctolagus
23	7	1.7	465	NP71_RAT	O62795 rattus norv
24	7	1.7	562	GCVK_HSV6U	P24446 herpes slmp
25	7	1.7	576	Y006_CAEEL	P34644 caenorhabdi
26	7	1.7	671	ALYS_ENTFA	P37710 enterococcu
27	7	1.7	835	Y422_MYCCE	P47661 mycoplasma
28	7	1.7	1597	R1R1_YEAST	P55552 saccharomyc
29	7	1.7	1804	YFAY_YEAST	P45883 saccharomyc
30	6	1.5	48	ATP8_SCHPO	P21336 schizosacch
31	6	1.5	49	LYSO_ECOLI	P02987 escherichia
32	6	1.5	60	MERC_PSEAE	P04139 pseudomonas
33	6	1.5	60	MERC_SHITL	P04337 shigella fl

34	6	1.5	64	1	RELX_RABER	P11952 raja erinac
35	6	1.5	101	1	CV02_CATRO	P25923 catharantnu
36	6	1.5	102	1	VE7_HPV36	P50811 human papil
37	6	1.5	102	1	YNI2_METTL	P05410 methanococc
38	6	1.5	103	1	VE7_HPV05	P06932 human papil
39	6	1.5	103	1	VE7_HPV08	P06430 human papil
40	6	1.5	103	1	VE7_HPV12	P36819 human papil
41	6	1.5	103	1	VE7_HPV47	P22423 human papil
42	6	1.5	103	1	VE7_HPV5B	P26559 human papil
43	6	1.5	108	1	IATP_PIG	O29307 sus scrofa
44	6	1.5	108	1	KVIV_HUMAN	P04430 homo sapien
45	6	1.5	115	1	TWC2_MOUSE	P03978 mus musculu

ALIGNMENTS

RESULT	ID	NP74_HUMAN	STANDARD:	PRT:	401 AA.
AC	000476:				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	15-JUL-1998 (Rel. 36, Last annotation update)				
DE	SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 4 (SODIUM/PHOSPHATE COTRANSPORTER 4) (NA(+)/PI COTRANSPORTER 4).				
DE	COTRANSPORTER 4) (NA(+)/PI COTRANSPORTER 4).				
GN	SLC17A4 OR NP74.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
RA	[1]				
RP	SEQUENCE FROM N.A.				
RA	Ruddy D.A., Kronmal G.S., Lee V.K., Muntler G.A., Quintana L., Domingo R., Meyer N.C., Basava A., McClelland E., Fullan A., Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchinashi Z., Wolff R.K., Schatzman R.C., Feder J.N.; Submitted (JUN-1997) to the EMBL/Genbank/DDJ databases.				
RL	-I- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).				
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.				
CC	-----				
DR	EMBL: U90545; AAB53423.1;				
KW	Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;				
KW	Sodium transport.				
FT	TRANSMEM 20				POTENTIAL.
FT	TRANSMEM 108				POTENTIAL.
FT	TRANSMEM 128				POTENTIAL.
FT	TRANSMEM 135				POTENTIAL.
FT	TRANSMEM 200				POTENTIAL.
FT	TRANSMEM 220				POTENTIAL.
FT	TRANSMEM 237				POTENTIAL.
FT	TRANSMEM 285				POTENTIAL.
FT	TRANSMEM 305				POTENTIAL.
FT	TRANSMEM 338				POTENTIAL.
FT	TRANSMEM 358				POTENTIAL.
FT	TRANSMEM 366				POTENTIAL.
FT	CARBOHYD 41				POTENTIAL.
FT	CARBOHYD 49				POTENTIAL.
FT	CARBOHYD 49				POTENTIAL.
FT	CARBOHYD 58				POTENTIAL.
FT	CARBOHYD 58				POTENTIAL.
SO	SEQUENCE 401 AA: 44051 MW: 5235982091948285 CRC64;				
Query Match 60.1%; Score 241; DB 1; Length 401;					
Best Local Similarity 100.0%; Pred. No. 4,7e+25;					
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	161	SWPMWTSKKEYIISSLKQVSSSKOPLPIKAMLSNPIMWISICIGCFSHQWVYVYI	220		
DB	161	STPMWTSKKEYIISSLKQVSSSKOPLPIKAMLSNPIMWISICIGCFSHQWVYVYI	220		

QY 221 PTYISSVYHVNIRDNGLLSALPEIYAVMGVGLADFLTKRKRLTVRKIATITLGL 280
| | | | |
DB 221 PTYISSVYHVNIRDNGLLSALPEIYAVMGVGLADFLTKRKRLTVRKIATITLGL 280
QY 281 PSSALIVSLPYINSGYITATALLTSCGLTSCSGIYINVDIAPRYSFFLMGASRGFS 340
| | | | |
DB 281 PSSALIVSLPYINSGYITATALLTSCGLTSCSGIYINVDIAPRYSFFLMGASRGFS 340
QY 341 STAPYIVPVSGFLSQDEPFGRVNFLLFVAVNLGLFLYILFEGADVOEWAKERKLT 400
| | | | |
DB 341 STAPYIVPVSGFLSQDEPFGRVNFLLFVAVNLGLFLYILFEGADVOEWAKERKLT 400
QY 401 L 401
|
DB 401 L 401

RESULT 2
NPT3_HUMAN STANDARD: PRT; 436 AA.
ID NPT3_HUMAN 000624;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 3 (SODIUM/PHOSPHATE
CORANSPORTER 3) (NA(+)/PI CORANSPORTER 3).
GN SLC17A3 OR NPT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA Ruddy D.A., Kromal G.S., Lee V.K., Mintier G.A., Quintana L.,
RA Domingo R., Meyer N.C., Basawa A., McClelland E., Fulan A.,
RA Mapa F.A., Moore T., Thomas W., Joeb D.B., Harmon C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
INTO CELLS VIA NA+ CORANSPORT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U91328; AAB82085.1; -;
DR EMBL; U90544; AAB53422.1; -;
KW Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
FT TRANSMEM 26
FT TRANSMEM 46
FT TRANSMEM 118
FT TRANSMEM 130
FT TRANSMEM 183
FT TRANSMEM 211
FT TRANSMEM 231
FT TRANSMEM 293
FT TRANSMEM 317
FT TRANSMEM 337
FT TRANSMEM 386
FT TRANSMEM 406
FT CARBOHYD 47
FT CARBOHYD 56
FT CARBOHYD 58
FT CARBOHYD 68
FT CARBOHYD 69
FT CARBOHYD 69
SQ SEQUENCE 436 AA; 47350 MM; DF02F618E83A572F CRC64;

Query Match 2.2%; Score 9; DB 1; Length 436;
Best Local Similarity 100.0%; Pred. No. 0.38;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 144 GCVCCLMP 152
| | | | |
DB 220 GCVCCLMP 228
RESULT 3
VE7_HP16 STANDARD: PRT; 98 AA.
ID VE7_HP16
AC P03129;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE E7 PROTEIN.
GN E7.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85246220.
RA Seedorf K., Kramer G., Durst M., Suhai S., Rowekamp W.G.;
RL "Human papillomavirus type 16 DNA sequence."; *J. Virol.* 145:181-185(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90218027.
RA Schneider-Maunoury S., Pehau-Arnaudet G., Breitburd F., Orth G.;
RL "Expression of the human papillomavirus type 16 genome in SK-V cells,
a line derived from a vulvar intraepithelial neoplasia."; *J. Gen. Virol.* 71:809-817(1990).
RN [3]
RP SEQUENCE FROM N.A.
RA Song Y.S., Kee S.H., Kim J.W., Park N.H., Kang S.B., Lee H.P.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Toranesello M.L., Buongiorno F.M., Meglio A., Buongiorno L.,
RA Beth-Girardo E., Girardo G.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
RN [5]
RP FUNCTION.
RX MEDLINE: 88223347.
RA Phelps W.C., Yee C.L., Munger K., Howley P.M.;
RL "The human papillomavirus type 16 E7 gene encodes transactivation and
transformation functions similar to those of adenovirus E1A."; *Cell* 53:539-547(1988).
CC -1- FUNCTION: E7 PROTEIN HAS BOTH TRANSFORMING AND TRANS-ACTIVATING
ACTIVITIES.
CC -1- DISEASE: HPV16, IN COMPARISON TO HPV TYPES 6 AND 11, IS MORE
OFTEN ASSOCIATED WITH MALIGNANT GENITAL CANCERS IN HUMANS.
CC -1- SIMILARITY: LOCAL WITH ADENOVIRUS E1A AND SV40 LT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; K02718; AAA46940.1; -;
DR EMBL; D00735; BAA00633.1; -;
DR EMBL; U76411; AAB18962.1; -;
DR EMBL; U76412; AAB18963.1; -;
DR EMBL; U76413; AAB18964.1; -;
DR EMBL; AF003020; AAB70737.1; -;
DR EMBL; AF003023; AAB70740.1; -;
DR EMBL; AF003024; AAB70741.1; -;
DR EMBL; AF003025; AAB70742.1; -;
DR EMBL; AF003026; AAB70743.1; -;
DR PTR; A03688; W7WLS.
DR PRAM; PF00527; E7; 1.
KW Early protein; Transcription regulation; Transforming protein;
DNA-binding; Trans-acting factor.

FT DOMAIN 58 61 C-XX-C MOTIF.
FT DOMAIN 91 94 C-XX-C MOTIF.
SQ SEQUENCE 98 AA; 11022 MW; 9BD612534CD2C9EB CRC64;

Query Match 1.7%; Score 7; DB 1; Length 98;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QLNDSSE 62
DB 27 QLNDSSE 33

RESULT 4
Y207_METJA
ID Y207_METJA STANDARD; PRT; 151 AA.
AC 057660;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN MJ0207.
GN MJ0207.
OS Methanococcus jannaschli.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 9633799.
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake O., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.N., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uitterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Bordovsky M.,
RA Klenk H.-P., Frazer C.M., Smith H.O., Moese C.R., Venter J.C.;
RA *Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschli.";
RT Science 273:1058-1073(1996).
RL -1- SIMILARITY: WEAK, TO M.JANNASCHII MJ1342 AND TO THE N-TERMINAL OF
M.JANNASCHII MJ1450.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D67476; AAB98194.1; -
DR TIGR; MJ0207; -
DR Hypothetical protein.
KW SEQUENCE 151 AA; 16912 MW; 03F78A3F2EBCA189 CRC64;
SQ

Query Match 1.7%; Score 7; DB 1; Length 151;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 125 GFISETL 131
DB 122 GFISETL 128

RESULT 5
RS5_CHLTR
ID RS5_CHLTR STANDARD; PRT; 165 AA.
AC P28543;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE 30S RIBOSOMAL PROTEIN S5.
GN RPSL OR RS5 OR CTS12.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-L2/434/BU;
RX MEDLINE; 92138612.
RA Kaul R., Gray G.J., Koehncke N.R., Gu L.J.;
RT "Cloning and sequence analysis of the Chlamydia trachomatis spc
RT ribosomal protein gene cluster".
RL J. Bacteriol. 174:1205-1212(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D/OW-3/CX;
RX MEDLINE; 99000809.
RA Stephens R.S., Kaiman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RL Science 282:754-759(1998).
CC -1- FUNCTION: PROTEIN S5 IS IMPORTANT IN THE ASSEMBLY AND FUNCTION OF
CC THE 30S RIBOSOMAL SUBUNIT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S5P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M80325; AAA23178.1; -
DR EMBL; AE001323; AAC68113.1; -
DR PIR; A45192; A45192.
DR HSSP; P02357; 1PKP.
DR PROSITE; PS00585; RIBOSOMAL_S5; 1.
DR PFAM; PF00333; Ribosomal_S5; 1.
KW Ribosomal protein.
SQ SEQUENCE 165 AA; 17762 MW; 66F3E0AA8409481A CRC64;

Query Match 1.7%; Score 7; DB 1; Length 165;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 301 ALTLSC 307
DB 146 ALTLSC 152

RESULT 6
Y271_SYNY3
ID Y271_SYNY3 STANDARD; PRT; 172 AA.
AC P73892;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 19.4 KD PROTEIN SLR0271 PRECURSOR.
GN SLR0271.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Cyanoecoccales; Synechocystis.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97061201.
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hikosawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Mureki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT *Synechocystis* sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90910; BAA17956.1; -
 KW Hypothetical protein; Signal.
 FT SIGNAL 1 24
 FT CHAIN 1 24
 FT SEQUENCE 172 AA: 19439 MW: 2606EB831BD440CA CRC64;
 SQ
 Query Match 1.7%; Score 7; DB 1; Length 172;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 299 ATALLTL 305
 Db 11 ATALLTL 17
 RESULT 7
 BACA_AZOBR STANDARD; PRT; 187 AA.
 AC P39438;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE BACTIRACIN RESISTANCE PROTEIN HOMOLOG (PUTATIVE UNDECAPRENOL KINASE)
 DE (EC 2.7.1.66) (FRAGMENT).
 OS *Azospirillum brasilense*.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
 OC *Azospirillum*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SP;
 RX MEDLINE; 93155143.
 RA Pellanda R., Vanoni M.A., Perego M., Pinbelli L., Galizzi A.,
 RA Curti B., Zanetti G.;
 RT "Glutamate synthase genes of the diazotroph *Azospirillum brasilense*.
 RT Cloning, sequencing, and analysis of functional domains.";
 RL J. Biol. Chem. 268:3099-3106(1993).
 RN [2]
 RP CONCEPTUAL TRANSLATION.
 RA Rudd K.E.;
 RL Unpublished observations (NOV-1994).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
 CC -1- SIMILARITY: TO E.COLI BACA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: L04300; -; NOT_ANNOTATED_CDS.
 KW Antibiotic resistance; Transmembrane; Transferase; Kinase.
 FT TRANSMEM 13 33
 FT TRANSMEM 45 65
 FT TRANSMEM 85 105
 FT TRANSMEM 108 128
 FT TRANSMEM 144 164
 FT NON_TER 187 187
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.

SQ SEQUENCE 187 AA: 19842 MW: E04C8995BDFDB63A CRC64;
 Query Match 1.7%; Score 7; DB 1; Length 187;
 Best Local Similarity 100.0%; Pred. No. 19;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 274 ATILGSL 280
 Db 169 ATILGSL 175
 RESULT 8
 ATPO_DROME STANDARD; PRT; 209 AA.
 ID ATPO_DROME
 AC Q24439;
 DT 15-FEB-2000 (Rel. 39, Created)
 DT 15-FEB-2000 (Rel. 39, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEIN PRECURSOR,
 DE MITOCHONDRIAL (EC 3.6.1.34) (OSCP).
 GN OSCP.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; *Drosophila*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Ovary;
 RX MEDLINE; 99168769.
 RA Caggese C., Ragone G., Perrini B., Moschetti R., De Pinto V.,
 RA Galizzi R., Barsanti P.;
 RT "Identification of nuclear genes encoding mitochondrial proteins:
 RT isolation of a collection of *D. melanogaster* cDNAs homologous to
 RT sequences in the Human Gene Index database.";
 RL Mol. Gen. Genet. 261:64-70(1999).
 CC -1- FUNCTION: THIS PROTEIN SEEMS TO BE PART OF THE STALK THAT LINKS
 CC CF(0) TO CF(1). IT EITHER TRANSMITS CONFORMATIONAL CHANGES FROM
 CC CF(0) INTO CF(1) OR IS IMPLICATED IN PROTON CONDUCTION.
 CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
 CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
 CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
 CC HAS THREE MAIN SUBUNITS: A, B AND C.
 CC -1- SIMILARITY: BELONGS TO THE ATPASE DELTA CHAIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X99666; CAA67980.1; -
 DR FLVBASE; FB94001691; OSCP.
 DR PROSITE; PS00389; ATPASE_DELTA; 1.
 DR PFAM; PF00213; OSCP; 1.
 KW Hydrolyase; ATP synthetase; CF(1); Hydrogen ion transport;
 KW Mitochondrion; Transist peptide.
 FT TRANSIT 1 209
 FT CHAIN 1 209
 FT SEQUENCE 209 AA: 22421 MW: 5004018524DFCAB2 CRC64;
 SQ
 Query Match 1.7%; Score 7; DB 1; Length 209;
 Best Local Similarity 100.0%; Pred. No. 21;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 373 VNILGL 379
 Db 102 VNILGL 108
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.
 FT POTENTIAL.

```

RESULT 9
VFPM_ECOLI
ID VFPM_ECOLI STANDARD: PRT: 212 AA.
AC P39295;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHEICAL 23.4 KD PROTEIN IN RNR-AIDB INTERGENIC REGION.
GN VFPM.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE: 95334362.
RA Buitand V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: U14003; AAA97081.1; -.
DR EMBL: AE000490; AAC77142.1; -.
DR ECGENE: EG12486; yjfm.
KW Hypothetical protein.
SQ SEQUENCE 212 AA; 23352 MW; 0C69332DA15AEAE4 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 212;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 350 VSGFLLS 356
    |||||
DB 131 VSGFLLS 137

RESULT 10
CB21_WHEAT
ID CB21_WHEAT STANDARD: PRT: 266 AA.
AC P04784;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DE 15-JUL-1999 (Rel. 38, Last annotation update)
DE CHLOROCHYL A-B BINDING PROTEIN PRECURSOR (LHCII TYPE I CAB) (LHCP).
GN WHA1.6.
OS Triticum aestivum (wheat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Triticum.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 85295969.
RA Lampka G.K., Morelli G., Chua N.-H.;
RT "Structure and developmental regulation of a wheat gene encoding the
RT major chlorophyll a/b-binding polypeptide.";
RL Mol. Cell. Biol. 5:1370-1378(1985).
-1- FUNCTION: THE LIGHT-HARVESTING COMPLEX (LHC) FUNCTIONS AS A LIGHT
RECEPTOR, IT CAPTURES & DELIVERS EXCITATION ENERGY TO PHOTOSYSTEMS
WITH WHICH IT IS CLOSELY ASSOCIATED. THE N-TERMINUS OF THE PROTEIN
EXTENDS INTO THE STROMA WHERE IT IS INVOLVED WITH ADHESION OF
CC GRANAL MEMBRANES AND PHOTOREGULATED BY REVERSIBLE PHOSPHORYLATION
CC OF ITS THREONINE RESIDUES; BOTH ARE BELIEVED TO MEDIATE THE

```

```

CC DISTRIBUTION OF EXCITATION ENERGY BETWEEN PHOTOSYSTEMS I AND II.
CC -1- SUBUNIT: THE LHC COMPLEX CONSISTS OF CHLOROPHYLLS (A & B) AND
CC CHLOROCHYL A-B BINDING PROTEINS.
CC -1- SUBCELLULAR LOCATION: CHLOROPLAST THYLAKOID MEMBRANE.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: M10144; AAA34260.1; -.
DR PIR: A23755; CDWT.
DR PFAM: PF00504; chloroa_b-bind; 1.
KW Chlorophyll; Photosynthesis; Photosystem I; Photosystem II;
KW Thylakoid membrane; Chloroplast; Transit peptide; Multigene family;
KW Transmembrane; Phosphorylation.
FT TRANSIT 1 33 CHLOROPLAST (PROBABLE).
FT CHAIN 34 266 CHLOROPHYLL A-B BINDING PROTEIN.
FT TRANSMEM 100 119 POTENTIAL.
FT TRANSMEM 152 172 POTENTIAL.
FT TRANSMEM 220 236 POTENTIAL.
SQ SEQUENCE 266 AA; 28264 MW; 9D5719759DF3C4E8 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 266;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 280 LPSSALI 286
    |||||
DB 21 LPSSALI 27

RESULT 11
UCRL_MAIZE
ID UCRL_MAIZE STANDARD: PRT: 273 AA.
AC P49727;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 01-OCT-1996 (Rel. 34, Last annotation update)
DE UBIOUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT, MITOCHONDRIAL
DE PRECURSOR (EC 1.10.2.2) (RIESKE IRON-SULFUR PROTEIN) (RISP).
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Zea.
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE: 92073358.
RA Huang J.T., Struck F., Matzinger D.F., Levings C.S. III;
RT "Functional analysis in yeast of cDNA coding for the mitochondrial
RT Rieske iron-sulfur protein of higher plants.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:10716-10720(1991).
-1- FUNCTION: COMPONENT OF THE UBIOUINOL-CYTOCHROME C REDUCTASE
COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
COUPLED TO ATP SYNTHESIS.
-1- CATALYTIC ACTIVITY: O2(2-) + 2 FERRICYTOCHROME C = O +
2 FERROCYTOCHROME C.
-1- SUBUNIT: BCI COMPLEX CONTAINS 10 SUBUNITS; 3 RESPIRATORY
SUBUNITS, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.
-1- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
-1- MISCELLANEOUS: THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S
PROTEIN.
-1- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
BACTERIAL, CHLOROPLAST).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: M77224; AAA33507.1; -
DR HSSP: P13372; 1RTE.
DR MAI2EDB; 30151; -
DR PRINTS; PR00162; RIESKE.
DR PROSITE; PS00199; RIESKE_1; 1.
DR PFMW; PF00355; RIESKE; 1.
KW Mitochondrion; Electron transport; Respiratory chain; Iron-sulfur;
KW Oxidoreductase; Inner membrane; Transmembrane; Transit peptide.
FT TRANSIT 1 61 MITOCHONDRION (POTENTIAL).
FT CHAIN 62 273 UBIQUINOL-CYTOCHROME C REDUCTASE IRON-
FT METAL 216 216 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 218 218 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 235 235 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 238 238 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT DISULFD 221 237 BY SIMILARITY.
SQ SEQUENCE 273 AA; 29835 MW; 43B08C16E9B1A2C CRC64;

Query Match 1.7%; Score 7; DB 1; Length 273;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 335 ASRGFS 341
|||||
Db 56 ASRGFS 62

RESULT 12
GCVK_HSV62 STANDARD; PRT; 280 AA.
AC P52446;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE POSSIBLE GANCICLOVIR KINASE (EC 2.7.1.1) (FRAGMENT).
OS U69 OR CH2R.
CN Herpes simplex virus (type 6 / strain 229).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Roseolovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 96195263.
RA Lindquester G.J., Inoue N., Allen R.D., Castelli J.W.,
RA Stamey F.R., Dambaugh T.R., O'Brian J.J., Danovich R.M.,
RA Frenkel N., Pellett P.E.;
RT "Restriction endonuclease mapping and molecular cloning of the human
RT herpesvirus 6 variant B strain 229 genome.";
RL Arch. Virol. 141:367-379(1996).
CC -1- FUNCTION: PHOSPHORYLATES THE ANTIVIRAL NUCLEOSIDE ANALOG
CC GANCICLOVIR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE HCMV UL97 / HVS 36 FAMILY.
CC -1- SIMILARITY: TO TYROSINE-PROTEIN KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: L14772; AAB06352.1; -
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE_NEG.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN KINASE_DOM; FALSE_NEG.
KW Transferase; Kinase; ATP-binding.

FT NON_TER 1 1 BY SIMILARITY.
FT ACT_SITE 31 31
SQ SEQUENCE 280 AA; 32348 MW; 930E2ED2A50A9107 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 375 LIGLIFY 381
|||||
Db 190 LIGLIFY 196

RESULT 13
SCRK_ECOLI STANDARD; PRT; 305 AA.
ID SCRK_ECOLI
AC P40713;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE FRUCTOKINASE (EC 2.7.1.4).
GN CSCK.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EC132;
RA Bockmann J.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + D-FRUCTOSE = ADP + D-FRUCTOSE
CC -1- SIMILARITY: BELONGS TO THE PFKB FAMILY OF CARBOHYDRATE KINASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: X81461; CA57218.1; -
DR PROSITE; PS00583; PFKB_KINASES_1; 1.
DR PROSITE; PS00584; PFKB_KINASES_2; 1.
DR PFMW; PF00294; PFKB; 1.
KW Transferase; Kinase.
SQ SEQUENCE 305 AA; 33081 MW; 5F9C8D59B66C3E55 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 305;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 105 LCSTALS 111
|||||
Db 125 LCSTALS 131

RESULT 14
LXB2_PHOUL STANDARD; PRT; 324 AA.
ID LXB2_PHOUL
AC P23147;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE ALKANAL MONOOXYGENASE BETA CHAIN (EC 1.14.14.3) (BACTERIAL LUCIFERASE
DE BETA CHAIN).
GN LUXB.
OS Photobacterium luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photobacterius.

```
RM [1]
RP SEQUENCE FROM N.A.
RC STRAIN-HW:
RX MEDLINE: 91139581.
RA "Cloning and nucleotide sequences of lux genes and characterization
RT of luciferase of Xenorhabdus luminescens from a human wound.";
RL J. Bacteriol. 173:1399-1405(1991).
CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE
CC SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY
CC REQUIRED FOR BIOLUMINESCENCE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: FMNH(2) + RCHO + O(2) -> FMN + RCOOH + H(2)O
CC + LIGHT.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M62917; AAA63566.1; -.
DR PIR: D38448; D38448.
DR HSSP: P07739; 1XKJ.
DR PRINTS: PR00089; LUCIFERASE.
DR PROSITE: PS00494; BACTERIAL_LUCIFERASE; 1.
DR PFAM: PF00296; bac_luciferase; 1.
DR Photoprotein; Luminescence; Oxidoreductase; Monoxygenase;
KW Flavoprotein; FMN.
SQ SEQUENCE 324 AA; 37115 MW; E7D81324D344A975 CRC64;

Query Match 1.7%; Score 7; DB 1; Length 324;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 349 TVSGFLL 355
DB 57 TVSGFLL 63

RESULT 15
LXBL_PHOTO STANDARD: PRT: 327 AA.
AC P19840:
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE ALKANAAL MONOOXYGENASE BETA CHAIN (EC 1.14.14.3) (BACTERIAL LUCIFERASE
DE BETA CHAIN).
GN LUXB.
OS Photorhabdus luminescens (Xenorhabdus luminescens).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Photorhabdus.
CC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-ATCC 29999;
RX MEDLINE: 90375532.
RA Sztitner R., Meighen E.;
RT "Nucleotide sequence, expression, and properties of luciferase coded
RT by lux genes from a terrestrial bacterium.";
RL J. Biol. Chem. 265:16581-16587(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 29999;
RX MEDLINE: 92355513.
RA Meighen E.A., Sztitner R.B.;
RT "Multiple repetitive elements and organization of the lux operons of
RT luminescent terrestrial bacteria.";
RL J. Bacteriol. 174:5371-5381(1992).
RN [3]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-HW:
RX MEDLINE: 90343746.
RA Johnston T.C., Rucker E.B., Cochrum L., Hruska K.S., Vandegrift V.;
RT "The nucleotide sequence of the luxA and luxB genes of Xenorhabdus
RT luminescens HM and a comparison of the amino acid sequences of
RT luciferases from four species of bioluminescent bacteria.";
RL Biochem. Biophys. Res. Commun. 170:407-415(1990).
CC -1- FUNCTION: LIGHT-EMITTING REACTION IN LUMINOUS BACTERIA. THE
CC SPECIFIC ROLE OF THE BETA SUBUNIT IS UNKNOWN, BUT IT IS ABSOLUTELY
CC REQUIRED FOR BIOLUMINESCENCE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: FMNH(2) + RCHO + O(2) -> FMN + RCOOH + H(2)O
CC + LIGHT.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M57416; AAA27624.1; -.
DR PIR: M55977; AAA27627.1; -.
DR HSSP: P07739; 1XKJ.
DR PRINTS: PR00089; LUCIFERASE.
DR PROSITE: PS00494; BACTERIAL_LUCIFERASE; 1.
DR PFAM: PF00296; bac_luciferase; 1.
DR Photoprotein; Luminescence; Oxidoreductase; Monoxygenase;
KW Flavoprotein; FMN.
KW VARIANT 85 A -> R (IN STRAIN HM).
SQ SEQUENCE 327 AA; 37595 MW; 1905AF01658BE56E CRC64;

Query Match 1.7%; Score 7; DB 1; Length 327;
Best Local Similarity 100.0%; Pred. No. 31;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 349 TVSGFLL 355
DB 57 TVSGFLL 63
```

Search completed: May 25, 2000, 14:59:49
Job time: 346 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:52:26 ; Search time 17.37 Seconds
(without alignments)
1600.632 Million cell updates/sec

Title: US-09-391-958-1
Perfect score: 401
Sequence: 1 MOVDELIPKRVPSICSARY.....LIFGEADVQEWAKRKILRL 401

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 225878 seqs, 69334122 residues

Word size : 0

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Listing first 45 summaries

Database :

SPTREMBL_12:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8	2.0	333	2 09Z6L1	09Z6L1 Chlamydia p
2	8	2.0	336	12 09WR9	09WR9 macaca mula
3	8	2.0	376	1 09TEG6	09TEG6 aetopyrum p
4	8	2.0	497	4 09Y2C5	09Y2C5 homo sapien
5	8	2.0	836	5 061343	061343 drosophila
6	7	1.7	49	9 038059	038059 bacterioph
7	7	1.7	63	4 015140	015140 homo sapien
8	7	1.7	68	10 008570	008570 catharanthu
9	7	1.7	97	5 019002	019002 caenorhabdi
10	7	1.7	98	12 012338	012338 human papil
11	7	1.7	128	1 058910	058910 pyrococcus
12	7	1.7	141	2 066368	066368 unidentified
13	7	1.7	142	2 P71192	P71192 escherichia
14	7	1.7	209	5 024439	024439 drosophila
15	7	1.7	235	2 09W253	09W253 thermocoga
16	7	1.7	248	8 09XLY0	09XLY0 hepiatulus ka
17	7	1.7	249	7 008607	008607 cyphotliapl
18	7	1.7	257	2 052995	052995 rhizobium m
19	7	1.7	264	5 097455	097455 caenorhabdi

20	7	1.7	283	2 006523	006523 haemophilus
21	7	1.7	283	2 087121	087121 actinobacil
22	7	1.7	283	2 09ZAB7	09ZAB7 actinobacil
23	7	1.7	296	5 021199	021199 caenorhabdi
24	7	1.7	299	2 088084	088084 enterococcu
25	7	1.7	304	13 091580	091580 xenopus lae
26	7	1.7	310	2 067396	067396 aquifex aeo
27	7	1.7	320	2 052533	052533 pseudomonas
28	7	1.7	324	2 056822	056822 xenorhabdus
29	7	1.7	344	4 000421	000421 homo sapien
30	7	1.7	344	6 09XSD7	09XSD7 macaca mula
31	7	1.7	356	4 075307	075307 homo sapien
32	7	1.7	480	2 054442	054442 serralia ma
33	7	1.7	480	2 09WYD3	09WYD3 serralia ma
34	7	1.7	484	4 013793	013793 homo sapien
35	7	1.7	492	2 P96820	P96820 mycobacteri
36	7	1.7	512	2 051750	051750 borrelia bu
37	7	1.7	512	3 013985	013985 schizosacch
38	7	1.7	512	10 082390	082390 arabidopsis
39	7	1.7	527	11 P70486	P70486 rattus norv
40	7	1.7	534	4 014611	014611 homo sapien
41	7	1.7	534	11 09WVK4	09WVK4 mus musculi
42	7	1.7	537	5 018274	018274 caenorhabdi
43	7	1.7	554	10 024298	024298 plisum sativ
44	7	1.7	554	10 065197	065197 plisum sativ
45	7	1.7	560	11 062634	062634 rattus norv

ALIGNMENTS

RESULT	1	ALIGNMENTS
ID	09Z6L1	PRELIMINARY; PRT; 333 AA.
AC	09Z6L1;	
DT	01-MAY-1999 (TREMBLrel. 10, Created)	
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT	01-MAY-1999 (TREMBLrel. 10, Last annotation update)	
DE	ASPARTATE DEHYDROGENASE.	
GN	ASD.	
OS	Chlamydia pneumoniae.	
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=CWL029;	
RA	KALMAN S., MITCHELL W., MARATHE R., LAMMEL C., FAN J., OLINGER L.,	
RA	GRIMWOOD J., DAVIS R.W., STEPHENS R.S.;	
RT	*Comparative Genomes of Chlamydia pneumoniae and C. trachomatis.*;	
RL	Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AE001685; AAD19185.1; -	
SQ	SEQUENCE 333 AA; 37300 MW; 0C9213A4 CRC32;	
Query Match	2.0%; Score 8; DB 2; Length 333;	
Best Local Similarity	100.0%; Pred. No. 8.4;	
Matches	8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
Qy	182 GSSKOPLP 189	
Db	206 GSSKOPLP 213	
RESULT	2	
ID	09WR9	PRELIMINARY; PRT; 336 AA.
AC	09WR9;	
DT	01-NOV-1999 (TREMBLrel. 12, Created)	
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)	
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)	
DE	HYPOPHOSPHATASE 36.9 KD PROTEIN.	
OS	Macaca mulatta rhadinovirus 17577.	
OC	Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Rhadinovirus.	

RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 99174001.
 RA SEARLES R.P., BERGOUAN E.P., AXTELM M.K., WONG S.W.:
 RT "Sequence and genomic analysis of a Rhesus macaque rhadinovirus with
 RT similarity to Kaposi's sarcoma-associated herpesvirus/human
 RT herpesvirus 8.";
 RL J. Virol. 73:3040-3053(1999).
 DR EMBL: AF083501; AAD21359.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 336 AA; 36913 MW; E016FACE CRC32;

Query Match	2.0%;	Score 8;	DB 12;	Length 336;
Best Local Similarity	100.0%;	Pred. No. 8.5;		
Matches	8;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0

QY	373	VNLTGLEF	380
Db	258	VNLTGLEF	265

RESULT	3
09YEG6	
ID	09YEG6
AC	09YEG6;
DT	01-NOV-1999 (TREMBLrel. 12, created)
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)
DT	01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE	376AA LONG HYPOTHEITICAL PROTEIN.
GN	AE06010.
OS	Aeropyrum pernix.
OC	Archaea; Crenarchaeota; Aeropyrum.
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=K1;
RX	MEDLINE; 99310339.
RA	KAMARABAYASI Y., HINO Y., HORIKAWA H., YAMAZAKI S., HAIKAWA Y., JIN-NO K., TAKAHASHI M., SEKINE M., BABA S., ANKAI A., KOSUGI H., HOSOIYAMA A., FUKUI S., NAGAI Y., NISHIJIMA K., NAKAZAWA H., TAKAKIYA M., MASUDA S., FUWABASHI T., TANAKA T., KUOHO Y., YAMAZAKI J., KUSHIDA N., OGUCHI A., AOKI K., KUBOTA K., NAKAMURA Y., NOMURA N., SANO Y., KIRKICI H.,
RT	"Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix K1.";
RT	Crenarchaeon, Aeropyrum pernix K1.";
RL	DNA Res. 6:83-101(1999).
DR	EMBL; AF000060; BAA79580.1;
SO	SEQUENCE 376 AA; 39239 MW; 775B0176 CRC32;

Query Match	2.0%;	Score 8;	DB 1;	Length 376;
Best Local Similarity	100.0%;	Pred. No. 9.3;		
Matches	8;	Conservative 0;	Mismatches 0;	Indels 0;
QY	299	ATATLTLS	306	
Db	26	ATATLTLS	33	

RESULT	4
09Y2C5	
ID	PRELIMINARY; PRT; 497 AA.
AC	09Y2C5;
DT	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, last sequence update)
DE	NA/P04 COTRANSPORTER HOMOLOG.
DE	Homo sapiens (Human)
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OC	[1]
RP	SEQUENCE FROM N.A.
RP	MEDLINE: 99253143.
XX	

RA SHIBUJI A., TSUNODA T., SEKI N., SUZUKI Y., SUGANE K., SUANO S.,
RT "Isolation and chromosomal mapping of a novel human gene showing
RT homology to Na⁺/P_o4 cotransporter.",
J. Hum. Genet. 44:190-192(1999).
DR EMBL: AB020527; BAA76663.1; -.
SQ SEQUENCE 497 AA; 54055 MW; 128F47B4 CRC32;

Query Match	2.0%	Score 8	DB 4	Length 497
Best Local Similarity	100.0%	Pred. No. 12		
Matches	8	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0

QY	360	EEGWRNYF	367
DB	456	EEGWRNYF	463

RESULT	5	
061343		
ID	061343	PRELIMINARY;
AC	061343.	PRT; 836 AA.

DT 01-Aug-1998 (TReMBLrel, 07, Created)
DT 01-Aug-1998 (TReMBLrel, 07, last sequence update)
DT 01-May-1999 (TReMBLrel, 10, last annotation update)
DE TWENTY.
GN TWENTY.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
IN [1]

RC STRAIN-CANTONS;
RX MEDLINE; 96129280.
RA DE COUET H.G., FONG K.S., WEEDS A.G., MC LAUGHLIN P.J., MIKLOS G.L.;
RT "Molecular and mutational analysis of a gelsolin-family member encoded
RT by the flightless I gene of *Drosophila melanogaster*.";
RL Genetics 141:1049-1059(1995).
121

RP SEQUENCE FROM N.A.
RC STRAIN-CANTONS;
RX MEDLINE; 97289742.
RA MIKILOS G.L., YAMAMOTO M., BURNS R.G., MALESZKA R.;
RT "An essential cell division gene of *Drosophila*, absent from
RT *Saccharomyces*, encodes an unusual protein with tubulin-like and
RT myosin-like peptide motifs.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:5189-5194(1997).
131

SEQUENCE FROM N.A.
RC STRAIN-CANTONS;
MALLESZKA R., DE COUET H.G., MIKLOS G.L.G.,
RA Proc. Natl. Acad. Sci. U.S.A. 0:0-0(1998).
DR ENBL: AF017777; AAC28400.1; -
SEQUENCE, FB00015588; tly.
DR PLEASE, 836 AA; 92838 MW; EE221871 CRC32
SO SEQUENCE

Query Match	2.0%;	Score 8;	DB 5;	Length 836;
Best Local Similarity	100.0%;	Pred. No. 18;		
Matches	8;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;

QY	375	LLGLLFYL	382
Db	63	LLGLLFYL	70

RESULT	6		
Q38059			
ID	Q38059	PRELIMINARY;	PRT; 49 AA.
ID	Q38059		
AC			
DT	01-NOV-1996	(TRENDArel 01, Created)	
DT	01-NOV-1996	(TRENDArel 01, Last sequence update)	
DT	01-NOV-1996	(TRENDArel 08, Last annotation update)	
DT	01-NOV-1998		


```

DE  HYPOTHETICAL 5.7 KD PROTEIN.
OS  Bacteriophage PZA.
OC  Viruses; dsDNA viruses, no RNA stage; Tailed phages; Podoviridae.
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE: 86056991.
RA  PACES V., VICEK C., URBANEK P., HOSTOMSKY Z.;
RT  "Nucleotide sequence of the major early region of Bacillus subtilis
   phage PZA, a close relative of phi 29."
RL  Gene 38:45-56(1985).
DR  EMBL: M1813; AAA8482.1; -.
KM  Hypothetical protein.
SQ  SEQUENCE 49 AA; 5650 MW; FD1BC640 CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 9; Length 49;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  315 SGITINV 321
    |||||
DB  29 SGITINV 35

RESULT 7
ID  015140 PRELIMINARY; PRT; 63 AA.
AC  015140:
DT  01-JAN-1998 (TREMBLrel. 05, Created)
DT  01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT  01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE  NA.K-ATPASE BETA-3 SUBUNIT PSEUDOGENE, COMPLETE SEQUENCE.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
   Mammalia; Primates; Catarrhini; Hominoidea; Homo.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  MALIK N., CAMPBELL V., SANCHEZ-WATTS G., WATTS A., SCHERR S.,
   BEATTY B., GROS P., LEVENSON R.;
RL  Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF005898; AAB61675.1; -.
DR  PFMW: PF00287; Na_K-ATPase; 1.
SQ  SEQUENCE 63 AA; 7591 MW; 4F685A8E CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 4; Length 63;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  378 LFFYLIF 384
    |||||
DB  37 LFFYLIF 43

RESULT 8
ID  008570 PRELIMINARY; PRT; 68 AA.
AC  008570:
DT  01-NOV-1996 (TREMBLrel. 01, Created)
DT  01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT  01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE  CYTOCHROME P-450 (EC 1.14.14.1) (CR9) (FRAGMENT).
GN  P450CR9.
OS  Catharanthus roseus (Rosy periwinkle) (Madagascar periwinkle).
OC  Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
   Euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
   core eudicots; Asteridae; euasterids I; Gentianales; Apocynaceae;
   Catharanthus.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  STRAIN(L.) G.DON;
RX  MEDLINE: 93283641.
RA  MEIJER A.H., SOUVER E., VERPOORTE R., HOGE J.H.C.;

```

```

RT  "Isolation of cytochrome P-450 cDNA clones from the higher plant
   Catharanthus roseus by a PCR strategy."
RL  Plant Mol. Biol. 22:379-383(1993).
CC  -1- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
   MONOOXYGENASES. MAY BE A GERANIOL-10-HYDROXYLASE.
CC  -1- CATALYTIC ACTIVITY: RH + REDUCED FLAVOPROTEIN + O(2) -> ROH +
   OXIDIZED FLAVOPROTEIN + H(2)O.
CC  -1- PATHWAY: INDOLE ALKALOID SYNTHESIS.
CC  -1- SUBCELLULAR LOCATION: MEMBRANE BOUND.
CC  -1- SIMILARITY: MEMBER OF THE CYTOCHROME P-450 FAMILY.
DR  EMBL: X69781; CAA49436.1; -.
DR  MENDEL: 11599; Ctr0:113; 11599.
KM  Oxidoreductase; Monooxygenase; Electron transport; Transmembrane;
   Heme.
FT  NON_TER 1 1
SQ  SEQUENCE 68 AA; 7579 MW; 752FBC06 CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 10; Length 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  175 SSLKQV 181
    |||||
DB  62 SSLKQV 68

RESULT 9
ID  019002 PRELIMINARY; PRT; 97 AA.
AC  019002:
DT  01-NOV-1996 (TREMBLrel. 01, Created)
DT  01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT  01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE  COSMID D2096.
GN  D2096.5.
OS  Caenorhabditis elegans.
OC  Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
   Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN  [1]
RP  SEQUENCE FROM N.A.
RA  WILSON R., AINSWORTH R., ANDERSON K., BAYNES C., BRKS M.,
   BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
   CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
   RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
   JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
   RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
   RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHONKKEEN R.,
   RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
   RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
   RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT  "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
   elegans."
RL  Nature 368:32-38(1994).
RN  [2]
RP  SEQUENCE FROM N.A.
RA  GEISEL C.;
RL  Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A.
RA  WATERSTON R.;
RL  Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR  EMBL: U40800; AAA81489.1; -.
SQ  SEQUENCE 97 AA; 11281 MW; E5A676F5 CRC32;

Query Match
Best Local Similarity 100.0%; Score 7; DB 5; Length 97;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  305 LSCGLST 311
    |||||

```

Db 34 LSCGLST 40

RESULT 10
ID 012338 PRELIMINARY; PRT; 98 AA.
AC 012338;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, last annotation update)
DE E7 PROTEIN.
OS Human papillomavirus type 16.
OC Viruses; dsDNA viruses, no RNA stage; Papovaviridae; Papillomavirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-PC-4;
RX MEDLINE; 97437474.
RA TORRESIELLO M.L., BUONAGURO F.M., MEGLIO A., BUONAGURO L.,
RA BERTI-GIRALDO E., GIRALDO G.,
RT "Sequence variations and viral genomic state of human papillomavirus
type 16 in penile carcinomas from Ugandan patients.";
RL J. Gen. Virol. 78:0-0(0).
DR EMBL; AF003022; AAB70739.1; -.
DR PFW; PF00527; E7; 1.
SQ SEQUENCE 98 AA; 10969 MW; 2D8CB74C CRC32;

Query Match 1.7%; Score 7; DB 12; Length 98;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 QUNDSE 62
Db 27 QUNDSE 33

RESULT 11
ID 058910 PRELIMINARY; PRT; 128 AA.
AC 058910;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, last annotation update)
DE 128AA LONG HYPOTHETICAL PROTEIN.
GN PH187.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE; 98344137.
RA KAWARABAYASHI Y., SAWADA M., HORIKAWA H., HAIKAWA Y., HINO Y.,
RA YAMAMOTO S., SETINE M., BABA S., KOSUGI H., HOSOMIYA A., NGAI Y.,
RA SAKAI M., OGURA K., OTUKA R., NAKAZAWA H., TAKAMURA M., OHYUKU Y.,
RA FUJIMASHI T., TANAKA T., KUDOH Y., YAMAZAKI J., KUSHIDA N., OGUCHI A.,
RA AOKI K., NAKAMURA Y., ROBB T.F., HORIKOSHI K., MASUCHI Y., SHIZUYA H.,
RA KIKUCHI H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000005; BAA30287.1; -.
SQ SEQUENCE 128 AA; 14403 MW; 0E4F59C0 CRC32;

Query Match 1.7%; Score 7; DB 1; Length 128;
Best Local Similarity 100.0%; Pred. No. 38;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 276 ILGLSPS 282
Db 116 ILGLSPS 122

RESULT 12

ID 066368 PRELIMINARY; PRT; 141 AA.
AC 066368;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, last annotation update)
DE ALPHA SUBUNIT OF DINITROGENASE REDUCTASE (FE PROTEIN) (FRAGMENT).
GN NIFH.
OS unidentified nitrogen-fixing bacteria.
OC Bacteria.
RN [1]
RP SEQUENCE FROM N.A.
RA OKUMURA M., NODA S., KUDO T.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB011958; BAA28491.1; -.
DR HSSP; P00456; 1CP2.
DR PROSITE; PS00692; NIFH_FRXC_2; 1.
DR PROSITE; PS00746; NIFH_FRXC_1; 1.
DR PFW; PF00142; fer4_NifH; 1.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 141 AA; 15075 MW; 100DF20C CRC32;

Query Match 1.7%; Score 7; DB 2; Length 141;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 140 FGVGCV 146
Db 64 FGVGCV 70

RESULT 13
ID P71192 PRELIMINARY; PRT; 142 AA.
AC P71192;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
DT 01-JAN-1999 (TREMBLrel. 09, last annotation update)
DE UPP30.5.
GN UPP30.5.
OS Escherichia coli.
OC Plasmid Incp-beta R751.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RX SMITH C.A., THOMAS C.M.;
RA "Comparison of the nucleotide sequences of the vegetative replication
RT origins of broad host range incp plasmids R751 and RK2 reveals
RT conserved features of probable functional importance.";
RL Nucleic Acids Res. 13:557-572(1985).
DR EMBL; U67194; AAC64458.1; -.
KW Plasmid.
SQ SEQUENCE 142 AA; 15017 MW; E95FED8A CRC32;

Query Match 1.7%; Score 7; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 42;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 217 VVYIPY 223
Db 136 VVYIPY 142

RESULT 14
ID 024439 PRELIMINARY; PRT; 209 AA.
AC 024439;
SQ SEQUENCE 141 AA; 15075 MW; 100DF20C CRC32;

```

DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE ATP SYNTHASE OLIGOMYCIN SENSITIVITY CONFERRAL PROTEIN PRECURSOR,
DE MITOCHONDRIAL (EC 3.6.1.34) (OSCP).
GN OSCP.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-OVARY;
RX MEDLINE: 99168769.
RA CAGGESH C., RAGONE G., PERRINI B., MOSCHETTI R., DE PINTO V.,
RA CAIZZI R., BARSANTI P.;
RT "Identification of nuclear genes encoding mitochondrial proteins:
RT isolation of a collection of D. melanogaster cDNAs homologous to
RT sequences in the Human Gene Index database.";
RL Mol. Gen. Genet. 261:64-70(1999).
CC -1- FUNCTION: THIS PROTEIN SEEMS TO BE PART OF THE STALK THAT LINKS
CC CF(0) TO CF(1). IT EITHER TRANSMITS CONFORMATIONAL CHANGES FROM
CC CF(0) INTO CF(1) OR IS IMPLICATED IN PROTON CONDUCTION.
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -1- SIMILARITY: BELONGS TO THE ATPASE DELTA CHAIN FAMILY.
DR EMBL: X99666; CAA67980.1; -.
DR FLYBASE: FBgn0016691; OSCP.
DR PROSITE: PS00389; ATPASE_DELTA. 1.
DR PRAM: PF00213; OSCP. 1.
DR PRINTS: PR00125; ATPASEDELTA.
RW Hydroxylase; ATP synthetase; CF(1); Hydrogen ion transport;
RW Mitochondrion; Transist peptide.
FT TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
FT CHAIN ? 209 ATP SYNTHASE OLIGOMYCIN SENSITIVITY
FT CONFERRAL PROTEIN.
SQ SEQUENCE 209 AA; 22421 MW; 444323B3 CRC32;

Query Match 1.7%; Score 7; DB 5; Length 209;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 373 VNLGL 379
DB 102 VNLGL 108

RESULT 15
O9W253
AC O9W253; PRELIMINARY; PRT; 235 AA.
ID O9W253;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN TM0584.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99287316.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFI D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RT "Evidence for lateral gene transfer between Archaea and bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA NELSON K.E., CLAYTON R.A., GILL S.R., GWINN M.L., DODSON R.J.,
RA HAFI D.H., HICKEY E.K., PETERSON J.D., NELSON W.C., KETCHUM K.A.,
RA MCDONALD L., UTTERBACK T.R., MALEK J.A., LINHER K.D., GARRETT M.M.,
RA STEWART A.M., COTTON M.D., PRATT M.S., PHILLIPS C.A., RICHARDSON D.,
RA HEIDELBERG J., SUTTON G.G., FLEISCHMANN R.D., WHITE O., SALZBERG S.L.,
RA SMITH H.O., VENTER J.C., FRASER C.M.;
RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AE001733; AAD35669.1; -.
SQ SEQUENCE 235 AA; 26394 MW; 50FA4BFA CRC32;

Query Match 1.7%; Score 7; DB 2; Length 235;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 110 LSGMLG 116
DB 180 LSGMLG 186

```

Search completed: May 25, 2000, 14:59:05
Job time: 399 sec

OM of: US-09-391-958-1 to: EST:* out_format: pfs
Date: May 25, 2000 3:10 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL=frimex.p2n.model -DEFY=xlp
-Q/cgr2_1/USFTO.spool/US09391958/rmtat_24052000_213605_8916/app-query.fasta.1
-DB=EST -QMT=fastap -SUFFIX=oligo.rst -GAPOP=4.500
-GAPOP=0.500 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-OGAPOP=0.500 -OGAPOP=0.050 -XGAPOP=60.000 -XGAPOP=60.000
-FGAPOP=6.000 -FGAPOP=7.000 -YGAPOP=60.000 -YGAPOP=60.000
-DELOP=6.000 -DELOP=7.000 -START=1 -MATRIX=oligo
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR_SCORE=quality
-THR_MIN=1 -ALIGN=15 -MODE=LOCAL -OUTFM=pfs -NORM=ext -MNLIN=0
-MAXLEN=1000000 -USER=US09391958 -ICPU=6 -ICPU=3 -NO_XLIPX -WAIT
-THREADS=1

Search information block:

Query: US-09-391-958-1
Query length: 401
Database: EST:*
Database sequences: 4857316
Database length: 202661650
Search time (sec): 602.260000

WARN: XGAPOP and YGAPOP must be equal. Assuming YGAPOP=XGAPOP=60.000
WARN: XGAPEXT and YGAPEXT must be equal. Assuming XGAPEXT=YGAPEXT=60.000

Score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_est20:AA858296	-	98.00	1729.05	609	AA858296 ob13b06.s1 NCI_CGAP_K1
gb_est24:AI922455	-	72.00	1271.07	243	AI922455 q921b09.x1 NCI_CGAP_K1
gb_est34:AI922430	-	66.00	1163.87	244	AI922430 q921b09.y5 NCI_CGAP_K1
gb_est24:AI922430	-	62.00	1091.06	297	AI922430 q921b09.x1 NCI_CGAP_K1
gb_est34:AI922430	-	61.00	1075.58	211	AI922430 q921b09.x5 NCI_CGAP_K1
gb_est34:AI922430	-	47.00	820.21	452	AI922430 q921b09.x3 NCI_CGAP_K1
gb_est18:AA076766	-	44.00	769.01	331	AA076766 z120d11.s1 Soares_feta
gb_est18:AA076766	-	16.00	273.21	173	AI247271 q160f11.x1 NCI_CGAP_K1
gb_est34:BI73160	-	11.00	176.57	455	BI73160 RRC111-11K11.PP RRC1-11
gb_est24:AI956449	-	10.00	159.31	454	AI956449 u169h08.x1 Sngano_mous
gb_est41:AV67452	-	9.00	145.07	270	AV67452 AV267452 RIKEN full-16
gb_est27:AI27938	-	9.00	141.48	452	AI27938 mm25d12.x1 Stratagene
gb_est12:AA276173	-	9.00	140.85	495	AA276173 vc31a07.x1 Barstead_MF
gb_gsa15:AO570823	-	9.00	140.79	478	AO570823 HS-5366.B1.F11.SP6E.RH
gb_est38:AA067082	-	9.00	140.46	49.88	AA067082 HS-5366.B1.F11.SP6E.RH
gb_gsa13:BI0499	-	9.00	135.36	95.92	BI0499 T6A18-SP6.2.TAMU Arabid
gb_est1:TI2786	-	8.00	130.86	160	TI2786 0024 Pyrococcus furiosus
gb_est20:AA907661	-	8.00	130.52	168	AA907661 om09e01.s1 Soares_NFL
gb_est18:AA010571	-	8.00	127.75	254	AA010571 z109c09.s1 Soares_feta
gb_est9:DI76900	-	8.00	127.29	270.10	DI76900 MUSE6B04 mouse embryonal
gb_est4:HA24687	-	8.00	126.86	278.23	HA24687 y141e06.s1 Soares_breast
gb_est40:AV228411	-	8.00	126.69	283	AV228411 AV228411 RIKEN full-16
gb_est44:AA183544	-	8.00	126.62	291	AA183544 xj75g12.x1 Soares_NFL
gb_est1:Z46183	-	8.00	126.62	294	Z46183 HSC25C11 normalized inf
gb_est46:HA359493	-	8.00	126.60	295	HA359493 46118 MARC 2P1G Sus sc
gb_est4:HA34546	-	8.00	126.48	299.79	HA34546 EST11574 Rat PC-12 cell
gb_est21:DA2663	-	8.00	126.41	303	DA2663 DA2663 Rice callus cDNA
gb_gsa1:CSN00DMH	-	8.00	126.23	309.61	AI091175 Arabidopsis thaliana g
gb_gsa3:BA4529	-	8.00	126.23	311	BA4529 HS-1059-B1-D02-MF.ab1.CT
gb_est9:AA090850	-	8.00	126.07	318	AA090850 yu0491.seq.F Human fet
gb_gsa3:BA48504	-	8.00	125.98	332	BA48504 RRC111-2E17.TV RRC1-11 H
gb_est9:AA116763	-	8.00	125.96	342	AA116763 mq35e05.r1 Barstead_MF
gb_est35:AI825450	-	8.00	125.38	351	AI825450 wb73b01.x1 NCI_CGAP_K1
gb_gsa3:BA04650	-	8.00	125.36	352	BA04650 cSRU-4099-u CSL flow sc
gb_est26:AI159583	-	8.00	125.17	362	AI159583 qy22d11.x1 NCI_CGAP_K1
gb_est29:AI606736	-	8.00	125.11	362	AI606736 vc15e05.y1 Barstead_MF
gb_est6:NA5217	-	8.00	125.17	365	NA5217 20257 Lambda-PRI2 Arabid
gb_gsa10:AO107667	-	8.00	125.07	367	AO107667 HS-3111.AI-H09.MR.CT
gb_est5:HA66779	-	8.00	125.03	369	HA66779 yf84g10.r1 Soares fetal

gb_est7:WA1381 - 8.00 125.00 362.55 371 WA1381 mc46e01.r1 Soares mou
gb_est17:CG68800 - 8.00 124.92 366.05 375 CG68800 CG68800.Yu1 Kohara un
gb_gsa11:AO191156 - 8.00 124.90 366.92 376 AO191156 HS-2195-B2_A05.MR C
gb_est9:AA104994 + 8.00 124.88 367.79 377 AA104994 ml50c05.x1 Stratagene
gb_est29:AI587324 - 8.00 124.74 374.77 385 AI587324 tq04h05.x1 NCI_CGAP
gb_est46:AA326686 + 8.00 124.74 374.77 385 AA326686 19580 MARC 2B0V Bos
seq_name: gb_est20:AA858296

seq_documentation_block:

LOCUS AA858296 609 bp mRNA 09-MAR-1998
DEFINITION ob13b06.s1 NCI_CGAP_K1d3 Homo sapiens cDNA clone IMAGE:1323539 3'
similar to TR:0004476 0004476 SODIUM PHOSPHATE TRANSPORTER. ; mRNA
sequence.

ACCESSION AA858296
VERSION AA858296.1 GI:2946598
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 609)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

COMMENT Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2285607.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbtp/image/image.html

FEATURES
Seq primer: -40m13 fwd. ET from Amersham
High quality sequence stop: 467.
Location/Qualifiers
1..609
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1323539"
/clone_id="NCI_CGAP_K1d3"
/lab_host="PH10B"

/note="organ: kidney; Vector: p773D-Pac (pharmacia) with
a modified polylinker; Site: 1: Not I; Site 2: Eco RI; 1st
strand cDNA was primed with a Not I - Oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified p773 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 166 a 118 c 114 g 211 t
ORIGIN

alignment_scores: 98.00 Length: 98
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AA858296/rev ..

Align seg 1/1 to reverse of: AA858296 from: 1 to: 609

304 Thrleusercysglyleuserthrleucysglinseryltyrlleas 320
|||||

603 ACGCTCTTCGGGATTAAGCACATTGTGTCAGTCAGGCAATTATATCA 554
 320 nValleuaspIleAlaProArGTYrSerSerPheLeuMetGlyAlaSerA 337
 |||||||
 553 TGTCTTAGATATGTGCTCAAGGATATTCAGTTTCTCATGGAGACATCA 504
 337 rGgIyPheSerSerIleAlaProValIleValProThrValSerGlyPhe 353
 |||||||
 503 GAGGATTTTCAGCATAGCACCGTCATTTGTACCCACTGTCAGCGGATTT 454
 354 LeuLeuSerGlnAspProGluPheGlyTPrParGAsnValPhePheLeu 370
 |||||||
 453 CTCTTAGTCAGGACCGCTGAGTTGGGTGGAGAGATCTCTTCTTCTGCT 404
 370 uPheAlaValAsnLeuLeuGlyLeuLeuPheTYrLeuIlePheGlyGlu 387
 |||||||
 403 GTTTGCCGTACCTGTTAGGACTACTCTTCACTCATATTTGGAGAAC 354
 387 lAspValGlnGluTPrPalalysGluArgLysLeuThrArgLeu 401
 |||||||
 353 CAGATGTCCAGAAATGGCTTAAGAGAAACCTCACTCGTTTA 310

seq_name: gb_est24:AI222455

seq_documentation_block:

LOCUS AI222455 243 bp mRNA EST 27-OCT-1998
 DEFINITION qg21b09.x1 NCI-CGAP Kid3 Homo sapiens cDNA clone IMAGE:1761689 3'
 similar to TR:000476 000476 SODIUM PHOSPHATE TRANSPORTER. ; mRNA
 sequence.
 AI222455
 VERSION AI222455.1 GI:3804658
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)
 On May 9, 1995 this sequence version replaced gi:803109.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.llnl.gov/bdrp/image/image.html

Seq primer: -40UP from Gibco
 High quality sequence stop: 164.

FEATURES

SOURCE

1. 243
 Location/Qualifiers

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1761689"
 /clone_1lb="NCI-CGAP_Kid3"
 /lab_host="DH10B"
 /note="Organ: Kidney; Vector: p773D-Pac (pharmacia) with
 a modified polylinker; Site1: Not I; Site2: Eco RI; 1st
 strand cDNA was primed with a Not I - oligo(dt) primer,
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not
 I and Eco RI sites of the modified p773 vector. mRNA
 source: 2 pooled kidneys. Library went through one round
 of normalization. Library constructed by Bento Soares and
 M. Fatima Bonaldo."
 75 a 57 c 50 g 61 t

BASE COUNT

ORIGIN

alignment_scores: 72.00 Length: 72
 Quality: 1.000 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x AI222455/rev ..

Align seg 1/1 to reverse of: AI222455 from: 1 to: 243

330 SerPheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaProValI 346
 |||||||
 243 AGTTTCTCATGTGGAGCATCAAGAGATTTTTCAGCATATACCTGCTCAT 194
 346 eValProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGly 363
 |||||||
 193 TGTACCCACTGTCAGCGGATTTCTTCTTAGTCAAGACCTGAGTTGGGT 144
 363 rParGAsnValPhePheLeuLeuPheAlaValAsnLeuLeuGlyLeu 379
 |||||||
 143 GGAGGATGTCTCTTCTGCTGTTGCCGTAACTGTTAGGACTACTC 94
 380 PheTYrLeuIlePheGlyGluAlaAspValGlnGluTPrPalalysGlu 396
 |||||||
 93 TTTTACCTCATATTTTGGAGAACGATGTCCAGAAATGGCTTAAGAGAG 44
 396 GlyLeuThrArgLeu 401
 |||||||
 43 AAAAATCACTCGTTTA 28

seq_name: gb_est34:AI792430

seq_documentation_block:

LOCUS AI792430 244 bp mRNA EST 02-JUL-1999
 DEFINITION qg21b09.y5 NCI CGAP Kid3 Homo sapiens cDNA clone IMAGE:1761689 5'
 similar to SW:NP74_HUMAN 000476 SODIUM-DEPENDENT PHOSPHATE
 TRANSPORT PROTEIN 4 ; mRNA sequence.
 AI792430
 VERSION AI792430.1 GI:5340146
 KEYWORDS EST.
 SOURCE human.

ORGANISM

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)
 On Jun 5, 1998 this sequence version replaced gi:3187571.
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.llnl.gov/bdrp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
 Original clone citation: see original entry for original citation
 information
 This 5' resequenced clone has no previous 5' data to verify this
 new read against
 Putative full length read
 The vector to vector length is 245
 Seq primer: -40RP from Gibco.

FEATURES
source
Location/Qualifiers
1. 244
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Kid3"
/lab_host="DH108"
/note="Organ: kidney; Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
61 a 49 c 53 g 79 t

ORIGIN

alignment_scores:
Quality: 66.00 Length: 66
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AT792430 ..

Align seg 1/1 to: AT792430 from: 1 to: 244

336 SerATGGLPheSerSerIleAlaProValIleValProThrValSerGI 352
|||||
19 TCAGAGAGATTTCGAGCATAGCACCTCTCATTTGACCCACATGTCAGCG 68
352 yPhleuLeuSerGIAspProGIuPheGIyTTPArGAsnValPhePheL 369
|||||
69 ATTCTCTTAGTCAGACCTGAGATTGGGTGAGAGAAATGCTTCTTCT 118
369 eulPheAlaValAsnLeuLeuGIyLeuLeuPheTyTLeuIlePheGI 385
|||||
119 TGGTGTTCGCTTAACCTGTAGACATACCTCTTACCTCATATTTGGA 168
386 GluAlaAspValGIuGIuTPAlaLysGIuArgLysLeuThrArgLeu 401
|||||
169 GAACGAGATGTCGAAGATGGGCTTAAGAGAGAAACTCATTCTTTA 216

seq_name: gb_est24:AI244777

seq_documentation_block:
LOCUS AI244777 297 bp mRNA EST 28-JAN-1999
DEFINITION g192f01.x1 NCI-CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1866937 3'
similar to SW:NP74_HUMAN O00476 SODIUM-DEPENDENT PHOSPHATE
TRANSPORT PROTEIN 4 ; , mRNA sequence.

ACCESSION AI244777
VERSION AI244777.1 GI:3840174
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 297)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced g1:2045043.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmett-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

JOURNAL
COMMENT

DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www.bio.liml.gov/bdrp/image/image.html

Insert Length: 707 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 156.

FEATURES
source
Location/Qualifiers
1. 297
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Kid3"
/lab_host="DH108"
/note="Organ: kidney; Vector: pRT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer, double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pRT73 vector. mRNA source: 2 pooled kidneys. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
94 a 68 c 59 g 75 t 1 others

ORIGIN

alignment_scores:
Quality: 62.00 Length: 62
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AI244777/rev ..

Align seg 1/1 to reverse of: AI244777 from: 1 to: 297

340 SerSerIleAlaProValIleValProThrValSerGIyPheLeuLeuSe 356
|||||
213 TCGACCATAGCACCTGTATTTGACCCATGACGATTTCTTCTTAG 164
356 rGlnAspProGIuPheGIyTTPArGAsnValPhePheLeuPheAlaV 373
|||||
163 TCAGAGACCTGAGATTGGGTGAGAGAAATGCTTCTTCTTGTTCGG 114
373 aLAsnLeuLeuGIyLeuLeuPheTyTLeuIlePheGIyGluAlaAspVal 389
|||||
113 TTAACCTGTAGGACTCTTCTTACCTCATATTTGGAGAACACATGTC 64
390 GlnGIuTPAlaLysGIuArgLysLeuThrArgLeu 401
|||||
63 CAAGATGGGCTTAAGAGAGAAACTCATTCTTTA 28

seq_name: gb_est32:AT733707

seq_documentation_block:
LOCUS AT733707 211 bp mRNA EST 14-JUN-1999
DEFINITION g921b09.x5 NCI-CGAP_Kid3 Homo sapiens cDNA clone IMAGE:1761689 3'
similar to SW:NP74_HUMAN O00476 SODIUM-DEPENDENT PHOSPHATE
TRANSPORT PROTEIN 4 ; , mRNA sequence.

ACCESSION AT733707
VERSION AT733707.1 GI:5054820
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 211)
NCI-MIDR-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)

JOURNAL

COMMENT On Dec 20, 1995 this sequence version replaced gi:1133358.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40UP from Glibco.
Location/Qualifiers
1. 211
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1761689"
/clone_11b="NCI-CGAP_K1d3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 64 a 49 c 45 g 53 t
ORIGIN

alignment_scores:
Quality: 61.00 Length: 61
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AT733707/rev ..

Align seg 1/1 to reverse of: AT733707 from: 1 to: 211

```
341 SerTlleAlaProValIleValProThrValSerGlyPheLeuLeuSerG1 357
|||||
210 ACCATGACACCTGTCATTTGACCTGACGCGATTTCTTCTAGTCA 161
|||||
357 naspprogiluphegllytrpargasnValPhePheLeuLeuPheAlaValA 374
|||||
160 GACACCTGAGTTGGGTGGAGAGATGCTCTTCTTGGTGGCGTTA 111
|||||
374 snLeuLeuGlyLeuLeuPheTyrrLeuLeuPheGlyGluAlaSPValGln 390
|||||
110 ACCGTGTAGACTACTCTTCACTCATATTGGAGAACAGATGTCGA 61
|||||
391 GluTrrPAlaLysGluArgLysLeuThrArgLeu 401
|||||
60 GAATGGGCTAAAGAGAGAAACTCACTCTTGA 28
|||||
```

seq_name: gb_est35:AT821963

seq_documentation_block:
LOCUS AT821963 452 bp mRNA EST 09-JUL-1999
DEFINITION OB13b06.x5 NCI-CGAP Kid3 homo sapiens cDNA clone IMAGE:1323539 3'
similar to SW:NP74_HUMAN 000476 SODIUM-DEPENDENT PHOSPHATE
TRANSPORT PROTEIN 4 ;, mRNA sequence.
ACCESSION AT821963

VERSION AT821963.1 GI:5441042
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 452)
TITLE NCI/NIH-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute / National Institute of Dental Research,
Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
Unpublished (1997)
JOURNAL On Mar 10, 1998 this sequence version replaced gi:2948996.
Other ESTs: ob13b06.y5
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/dbp/image/image.html

This read is a RESEQUENCE of a previously sequenced human clone
Original clone citation: National Cancer Institute, Cancer Genome
Anatomy Project (CGAP), Tumor Gene Index
This read has been verified (found to hit its original self in the
correct orientation)
Seq primer: -40UP from Glibco.
Location/Qualifiers
1. 452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1323539"
/clone_11b="NCI-CGAP_K1d3"
/lab_host="DH10B"
/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer,
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not
I and Eco RI sites of the modified pT73 vector. mRNA
source: 2 pooled kidneys. Library went through one round
of normalization. Library constructed by Bento Soares and
M. Fatima Bonaldo."

BASE COUNT 115 a 84 c 78 g 175 t
ORIGIN

alignment_scores:
Quality: 47.00 Length: 47
Ratio: 1.000 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x AT821963/rev ..

Align seg 1/1 to reverse of: AT821963 from: 1 to: 452

```
355 LeuSerGlnAspProGluPheGlyTrpArgAsnValPhePheLeuLeuPh 371
|||||
450 CTTAGTCAGAGACCCCTGAGCTTTGGGTGGAGAGATGCTTCTTGGTGT 401
|||||
371 eAlaValAsnLeuLeuGlyLeuLeuPheTyrrLeuLeuPheGlyGluAlaA 388
|||||
400 TGGCGTTAAGCTGTTAGACTACTCTTCACTCATATTGGAGAACGCG 351
|||||
388 SPValGlnGluTrrPAlaLysGluArgLysLeuThrArgLeu 401
|||||
350 ATGTCGCAAGATGGGCTAAAGAGAGAGAAACTCACTCTTGA 310
|||||
```


IMAGE:1887711 3', mRNA sequence.

DEFINITION AV267452 RIKEN full-length enriched, adult male testis (DH10B) Mus

ACCESSION	musculus cDNA clone 4930526G11 3', mRNA sequence.
VERSION	AY267452
KEYWORDS	AY267452.1 GI:6255489
SOURCE	EST.
ORGANISM	house mouse. Mus musculus
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus. 1 (bases 1 to 270)
AUTHORS	Kono,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Ishikunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Horii,F., Ikeda,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,T., Kikuchi,N., Koima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugihara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateno,M., Tomiなが,N., Tsunoda,T., Watabiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y. RIKEN Mouse ESTs (Kono,H., et al.) Unpublished (1999) On May 14, 1999 this sequence version replaced gi:4827795. Contact: Yoshihide Hayashizaki Genome Exploration Research Group, Life Science Tsukuba Center, Genome Science Laboratory The Institute of Physical and Chemical Research (RIKEN), Genomic Sciences Center 3-1-1 Koyada, Tsukuba, Ibaraki 305-0074, Japan Tel : +81-298-36-9013 Fax: +81-298-36-9098 Email: genome-res@rtc.riken.go.jp/ URL:http://genome.rtc.riken.go.jp/ Sasaki,N., Izawa,M., Watabiki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998) Itoh,M., Kitsuana,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci,P. and Hayashizaki,Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999) Please visit our web site (http://genome.rtc.riken.go.jp) for further details.
FEATURES	Location/Qualifiers
SOURCE	1..270 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="4930526G11" /clone_lib="RIKEN full-length enriched, adult male testis (DHIOB)" /sex="male" /tissue_type="testis" /dev_stage="adult" /lab_host="DHIOB" /note="Site_1: Sali; Site_2: BamHI. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGATCCAGACTCTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using reversease thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGATTCGACTTAAATTAAATTAATCCCCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a

```

modified pBluescript KS(+ ) after bulk excision from Lambda
FLC I: Cloning sites, 5' end: Salt, 3' end: BamHI."
BASE COUNT      72 a      34 c      74 g      70 t
ORIGIN
alignment_scores:
    Quality:      9.00      Length:      9
    Ratio:      1.000      Gaps:      0
    Percent similarity: 100.000      Percent identity: 100.000
alignment_block:
US-09-391-958-1 x AV267452/rev ..
Align seg 1/1 to reverse of: AV267452 from: 1 to: 270
376 LeuGlyLeuLeuPheTyLeuIlePhe 384
|||||
212 TTGGGCGCTCCTGTTTACTTAATATT 186
seq_name: gb_est27:AI427938
seq_documentation_block:
LOCUS      AI427938      452 bp      mRNA      EST      09-MAR-1999
DEFINITION      mm25d12.x1 Stratiogene mouse skin (#937313) Mus musculus cDNA clone
TRANSGENE:522551 3' similar to TR:000476 000476 SODIUM PHOSPHATE
TRANSPORTER. ; mRNA sequence.
AI427938
AI427938.1 GI:4273864
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 452)
Marra,M., Hillier,L., Kueba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Persson,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurr,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The WashU-NCI Mouse EST Project 1999
Unpublished (1999)
On Feb 17, 1998 this sequence version replaced gi:2889535.
Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LMLT ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 400.
Location/Qualifiers
1. .452
/oranism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone IMAGE:522551"
/clone_id="Stratiogene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SJR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5'
adaptor sequence: 5' GAATTCGCGCACG 3'-3' adaptor
sequence: 5' CTCGAGTTTATTTTATTTT 3'"
BASE COUNT      113 a      119 c      79 g      140 t      1 others

```

124 a	129 c	88 g	154 t
-------	-------	------	-------

alignment_block:

alignment_block:

Mon Jun 12 23:02:44 2000

US-09-391-958-1 x A0570823 ..

Align seg 1/1 to: A0570823 from: 1 to: 499

302 LeuLeuThleuSerCysGlyLeuSer 310
 ||||||||||||||||||
 458 CTGTTGACTCTCTCGTGGTATCC 484

seq_name: gb_est38:AM067082

seq_documentation_block:

LOCUS AM067082 523 bp mRNA EST 12-OCT-1999
 DEFINITION 683015C02.x1 683 - 14 day immature embryo from Hake lab (HS) Zea
 mays cDNA, mRNA sequence.

ACCESSION AM067082
 VERSION AM067082.1 GI:6022154

KEYWORDS EST.
 SOURCE Zea mays.

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Zea.
 1 (bases 1 to 523)

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1999)
 On Dec 20, 1995 this sequence version replaced gi:1135638.
 Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 723 8221
 Email: walbot@stanford.edu
 Plate: 683015 row: C column: 02.

FEATURES

source

1..523
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="683 - 14 day immature embryo from Hake lab
 (HS)"
 /tissue_type="embryo"
 /dev_stage="14 days after pollination"
 /lab_host="DH10B"
 /note="Organ: embryo; Vector: PBKCMV (Stratagene's Zap
 Express); Site_1: XhoI; Site_2: EcoRI; directionally
 cloned, 14 day immature embryo library created with
 Stratagene's Zap Express cDNA protocol."

BASE COUNT 92 a 160 c 177 g 94 t

ORIGIN

alignment_scores:

Quality:	9.00	Length:	9
Ratio:	1.000	Gaps:	0
Percent Similarity:	100.000	Percent Identity:	100.000

alignment_block:

US-09-391-958-1 x AM067082 ..

Align seg 1/1 to: AM067082 from: 1 to: 523

277 LeuGlySerLeuProSerSerAlaLeu 285
 ||||||||||||||||||

78 CTCGATCTCTCTCCCTCCGCTTA 104

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 13:32:03 ; Search time 28.66 Seconds

(without alignments)
331.407 Million cell updates/sec

Title: US-09-391-958-1

Perfect score: 2076
Sequence: 1 MOVDETLIPRKVPSLCSARY.....LIFGEADVQIMAKERKUTRL 401Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2076	100.0	401	1 W69971	Human sodium-depen
2	2061	99.3	401	1 W78920	Human haemochromat
3	744	35.8	436	1 W78919	Human haemochromat
4	479	23.1	560	1 W05148	Human brain sodium
5	479	23.1	560	1 W70500	Human sodium-11thi
6	466.5	22.5	576	1 W88523	Eat-4 protein amin
7	169.5	8.2	482	1 W14439	Protein involved i
8	122.5	5.9	462	1 W97820	Staphylococcus aur
9	112.5	5.4	2969	1 W56446	Fragment HGJ2141 o
10	111.5	5.4	2965	1 W56450	Fragment F1029 of
11	111.5	5.4	2970	1 W56445	Fragment HGJ1916 o
12	110.5	5.3	2873	1 W56441	Fragment HGU606 of
13	110.5	5.3	2969	1 W56443	Fragment HGJ1741 o
14	108.5	5.2	3011	1 R95020	Hepatitis GB virus
15	107.5	5.2	470	1 R99353	Human amine transp
16	107.5	5.2	2860	1 W63611	Hepatitis GB virus
17	107.5	5.2	2905	1 W06536	Hepatitis GB virus
18	107.5	5.2	2969	1 W56442	Fragment HGJ1737 o
19	106.5	5.1	2873	1 R87559	Hepatitis virus cl
20	106.5	5.1	2873	1 R90796	HGV-PNF 2161 polyp
21	106.5	5.1	2873	1 W34983	Hepatitis G virus
22	106.5	5.1	2873	1 W76054	HGV isolate PNF 21
23	106.5	5.1	2873	1 W80148	HGV-PNF2161 varian
24	106.5	5.1	2873	1 W89452	Hepatitis G virus
25	106.5	5.1	2873	1 W92755	US856134 Seq ID 1
26	106.5	5.1	2969	1 W56449	Fragment K3732(2v)
27	104.5	5.0	470	1 W77493	Human amine transp
28	104.5	5.0	2910	1 R87566	Hepatitis G virus
29	104.5	5.0	2910	1 R90797	Hepatitis G virus
30	104.5	5.0	2910	1 W76088	HGV-JC variant pol
31	104.5	5.0	2910	1 W80182	Hepatitis G virus
32	104.5	5.0	2910	1 W89458	Hepatitis G virus
33	104.5	5.0	2910	1 W92789	US856134 Seq ID 1
34	104	5.0	502	1 W99600	B.stationis multi-

35	103.5	5.0	488	1 W37990	Mutant Aspergillus
36	103.5	5.0	514	1 R47342	Mammalian synaptic
37	103.5	5.0	514	1 W38286	Human synaptic ves
38	103.5	5.0	561	1 W14440	Protein Involved i
39	103.5	5.0	2723	1 W56448	Fragment HGJ1775 o
40	103	5.0	606	1 W35808	Cercospora kikuchi
41	102.5	4.9	559	1 W64554	Human liver cell c
42	102.5	4.9	2841	1 W06537	Hepatitis GB virus
43	101.5	4.9	537	1 W41195	Mouse osteoclast t
44	99.5	4.8	2969	1 W56447	Fragment HGJ168 o
45	99	4.8	456	1 P91895	Protein sequence o

ALIGNMENTS

RESULT	1	
ID	W69971	
AC	W69971: standard; Protein: 401 AA.	
DF	30-NOV-1998 (first entry)	
DE	Human sodium-dependent phosphate cotransporter.	
KW	Sodium-dependent phosphate cotransporter; human; NAFPR; cancer; myopathy;	
KW	cell signalling disorder; phosphate regulation disorder; therapy.	
OS	Homo sapiens.	
PN	W09837198-AL.	
PD	27-AUG-1998.	
PF	24-FEB-1998; U03745.	
PR	24-FEB-1997; US-805118.	
PA	(INCY-) INCYTE PHARM INC.	
PI	Bandman O, Lal P;	
DR	WPT: 98-467561/40.	
DR	N-PSDB: V43711.	
PT	New isolated human sodium-dependent phosphate co:transporter - used	
PT	to develop products for treating e.g. cancers, osteoporosis,	
PT	Alzheimer's disease, diabetes, encephalopathy, myopathy,	
PT	hypocalciuria or hypoglycaemia.	
PS	Claim 1; Fig 1; 66pp; English.	
CC	This sequence is the human sodium-dependent phosphate cotransporter	
CC	(NAFPR) of the invention. NAFPR and agonists of it can be used to treat	
CC	or prevent disorders associated with decreased phosphate levels,	
CC	e.g. cancers of the kidney, disorders of decreased phosphate levels	
CC	including tumoral calcinosis, osteomalacia, osteoporosis, familial	
CC	hypophosphataemia, rickets, cystinuria, nephrocalcinosis,	
CC	glomerulonephritis, renal calculus, Alzheimer's disease, diabetes	
CC	mellitus, hereditary amyloidosis, myopathies including progressive	
CC	external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic epilepsy,	
CC	encephalopathy, cardiomyopathy, hypokalaemia, Goodpasture syndrome, and	
CC	disorders of cell signalling through cAMP, ATP, NADPH and	
CC	glucose-6-phosphate. Antagonists or inhibitors of NAFPR may be	
CC	administered to a subject to treat or prevent disorders associated with	
CC	increased phosphate levels, e.g. hypocalciuria, hypocalcaemia, and	
CC	abnormal phosphate regulation in neurons, gastrointestinal tract and	
CC	liver. The products can also be used for detection, diagnosis and drug	
CC	screening. 401 AA:	
SQ	Sequence	
Query Match	100.0%; Score 2076; DB 1; Length 401;	
Best Local Similarity	100.0%; Pred. No. 3.3e-213;	
Matches	401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY	1 MOVDETLIPRKVPSLCSARYGALVLFPCNFTTIAONVIMNTVMVNVSTSPQSLNDS 60	
DB	1 MOVDETLIPRKVPSLCSARYGALVLFPCNFTTIAONVIMNTVMVNVSTSPQSLNDS 60	
QY	61 SEVLPVDSRGSKAPKSLPAKSSITGGOPATWERGPPQESRLCSIALSGMLGCFPA 120	
DB	61 SEVLPVDSRGSKAPKSLPAKSSITGGOPATWERGPPQESRLCSIALSGMLGCFPA 120	
QY	121 IILGGFISETLMPFVFYIFGGGVCCLMPFVIVDDVSPVWISTSEKRYTISLKKQ 180	
DB	121 IILGGFISETLMPFVFYIFGGGVCCLMPFVIVDDVSPVWISTSEKRYTISLKKQ 180	

```

OY 181 VSSKQPLPIKAMLSLPIWISICLCFSGHQLVSTWVYIPIYISSYHVNRDGLISA 240
DB 181 VSSKQPLPIKAMLSLPIWISICLCFSGHQLVSTWVYIPIYISSYHVNRDGLISA 240
OY 241 LPIYAWYIGWGVGLADFLITKFKRLITVRKIATILGSLPSALIVSLPYNSGYTAT 300
DB 241 LPIYAWYIGWGVGLADFLITKFKRLITVRKIATILGSLPSALIVSLPYNSGYTAT 300
OY 301 ALLTSCGLSTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
DB 301 ALLTSCGLSTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
OY 361 FGMWVFFLLFAVNLGLLFYLIIEGADVQEWAKERKLTRL 401
DB 361 FGMWVFFLLFAVNLGLLFYLIIEGADVQEWAKERKLTRL 401

RESULT 2
W78920
ID W78920 standard; Protein; 401 AA.
AC W78920:
DE 21-DEC-1998 (first entry)
DE Human haemochromatosis protein NTP4.
KW Bovine butyrophillin; BR; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BRP1; BRP2; BRP3;
KW BRP4; BRP5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene.
OS Homo sapiens.
PN MO9814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
DR WPI: 98-240014/21.
DR N-PSDB: V57910.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 60; Fig 5b; 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BRP genes, which are homologues of the milk
CC protein butyrophillin (BRP), and can be used in the production of agonists
CC and antagonists of BR function. Also described are: (1) a Roret gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia. The present sequence represents
CC NTP4.
SQ Sequence 401 AA;

Query Match 99.3%; Score 2061; DB 1; Length 401;
Best Local Similarity 99.3%; Pred. No. 1.3e-211;
Matches 398; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

```

DB 61 SEVLVDVSFGSLKAPKSLPAKSSILGQFAIWERWGPQERSRLCSIALSGMLIGCFTA 120
OY 121 ILIGFISFETLQMPFVFYIFGGVCCCLMPFVYIDDPVSYPWISTEKEKTIISLKOQ 180
DB 121 ILIGFISFETLQMPFVFYIFGGVCCCLMPFVYIDDPVSYPWISTEKEKTIISLKOQ 180
OY 181 VSSKQPLPIKAMLSLPIWISICLCFSGHQLVSTWVYIPIYISSYHVNRDGLISA 240
DB 181 VSSKQPLPIKAMLSLPIWISICLCFSGHQLVSTWVYIPIYISSYHVNRDGLISA 240
OY 241 LPIYAWYIGWGVGLADFLITKFKRLITVRKIATILGSLPSALIVSLPYNSGYTAT 300
DB 241 LPIYAWYIGWGVGLADFLITKFKRLITVRKIATILGSLPSALIVSLPYNSGYTAT 300
OY 301 ALLTSCGLSTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
DB 301 ALLTSCGLSTLCOSGIYINVDIAPRYSSFLMGASRGSSIAPIYPTVSGFLLSQDE 360
OY 361 FGMWVFFLLFAVNLGLLFYLIIEGADVQEWAKERKLTRL 401
DB 361 FGMWVFFLLFAVNLGLLFYLIIEGADVQEWAKERKLTRL 401

RESULT 3
W78919
ID W78919 standard; Protein; 436 AA.
AC W78919:
DE 21-DEC-1998 (first entry)
DE Human haemochromatosis protein NTP3.
KW Bovine butyrophillin; BR; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BRP1; BRP2; BRP3;
KW BRP4; BRP5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene.
OS Homo sapiens.
PN MO9814466-A1.
PD 09-APR-1998.
PF 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
DR WPI: 98-240014/21.
DR N-PSDB: V57909.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 54; Fig 5b; 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BRP genes, which are homologues of the milk
CC protein butyrophillin (BRP), and can be used in the production of agonists
CC and antagonists of BR function. Also described are: (1) a Roret gene
CC which can be used to develop products for the study, diagnosis and
CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
CC which are homologues of a type 1 sodium transport gene, and can
CC similarly be used for hypophosphatemia. The present sequence represents
CC NTP3.
SQ Sequence 436 AA;

Query Match 35.8%; Score 744; DB 1; Length 436;
Best Local Similarity 38.5%; Pred. No. 4.5e-71;
Matches 162; Conservative 56; Mismatches 107; Indels 96; Gaps 5;

```


Oy		10	RKVPSCARIGALVLFHCFETIIAQNVIINNTWAMANSTSPQSOINDSEYLPVDST	69
Dy		8	RKGDPGCLRYRGGLALIMHFNSFTMITORVSLSIAMIAMVTQOQGSLNSTEGPVADAF	67
Oy		70	GGLSKAPPSLPARKSII-----LG----	88
Dy		68	NNSSISIKFEPTRKASYQWSPETOGIIFSSINGIILITIPSGYLAGIFCAKKMLGAGLL	127
Oy		88	-----GFPIIWERGWGPQERSRLCSIALS	111
Dy		128	ISLTLTPTPLAAPOGVLYIVMTVTQGMQMGMATMTOGITIMAKMPRLPERSKITITAGS	187
Oy		112	GMLACGTFAIIIGERISELTGMPFEVYIFGGVCVCLLMFYVIYDDPVSPWLSTSEKE	171
Dy		188	GSACFSFIILCVGLISQALSMPEIFEFYGSTGCVCCLLMFTVYIDDDPHHPICSVREKE	247
Oy		172	XIISIMAQOVSSSNQPLPIKMKMLSLPLMSICLCGEFHQLVYSIMVYVFETYISSYTHVN	231
Dy		248	HILSSLAOOPSSPPRPAVKPMKMWCLPLMALEFLGFESHFMCLTIILTYPLTYISTLHVN	307
Oy		232	IRDNGLISALEFIFYAWVYIGNVGVLADPLFKTK-FRLITVRKIATI-----L	277
Dy		308	IRDSSVLSLPTFLIAASTTIIGGLADLPLSRNLRIILTPRKPLESSLDMDVSSMESOGDL	367
Oy		278	GSLPSALIVSLPYLINSGYIRATALLTSCGLSTPLCOSGIYINWLDIAPRVSSFMLGNASR	337
Dy		368	GSSQESSLPRLDSSSVARIILSVCGMFSFCULOSTCLANFTSRLD----KQNFKTGPKR	423
Oy		338	G 338	
Dy		424	G 424	
RESULT	4			
ID	W05148			
AC	W05148:			
DT	30-JAN-1997	(first entry)		
DE	Human brain sodium-dependent inorganic phosphate cotransporter.			
KW	Brain sodium-dependent inorganic phosphate cotransporter; hBNPI.			
OS	Homo sapiens.			
PN	MO9634288-A1.			
PD	31-OCT-1996.			
PE	25-APR-1996; U05792.			
PR	27-APR-1995; US-430033.			
PA	(ELIL) LILLY & CO ELI.			
PI	Nl B, Paul SM;			
DR	WPI; 96-497773/49.			
NR	N-PDB: T42064.			
PT	New isolated human brain sodium-dependent inorganic phosphate			
PT	co-transporter - used to develop prods for diagnosis, treatment and			
PS	prevention of conditions involving inappropriate stimulation			
CC	Claim 1; Page 55-57; 68pp; English.			
CC	A novel human brain sodium-dependent inorganic phosphate			
CC	cotransporter (W05148), designated hBNPI, is selectively expressed			
CC	in discrete populations of neurons and glia. Its amino acid			
CC	sequence was deduced from a cDNA clone (T42064) isolated from a			
CC	human hippocampus cDNA library. hBNPI can be expressed in			
CC	transformed host cells and used to determine the effectiveness			
CC	of test cpds. for the treatment or prevention of disorders			
CC	associated with an inappropriate stimulation of hBNPI.			
Sequence	560 AA;			

Query Match	23.1%	Score 479;	DB 1;	Length 560;
Best Local Similarity	28.2%	Pred. NO. 1.2e-42;		
Matches	112;	Conservative	68;	Mismatches 135; Indels 62; Gaps 7.
Oy	12	VPSLCASRGALVLMHFCFETTTIAONTINMTVMAMVNSTSPQSOLNDSSEVLVYDVSFGG	71	
db	161	IPSAARVHYGCYI-----FVRITLGGILEGCVTPPC-----	191	

```

07 72 LSNAPKSLPAWSSILGGGFALWEMWGPQDERSRLCSIALSGMLGCTALLIGFISETL 131
Db 191 -----HGISWAPLERSRIATTAFCGSYAGAVAPMLAGLVQYS 232
07 132 GWPEFYIFGCVGCCCLMFVVIYDDPVSPWISTSEKEITYISLKOQVSSKO---PL 188
Db 233 GMSSTFYVYGSFGIFWYLFMLVLTYESALHPISISEEKRY----EDALGESAKMLNPL 288
07 189 -----PIKAMLRSLDPWISICLGCTSHQWLVTWVVIPTIYISSYVHVNIRDNGLSALPF 243
Db 289 TKFSTPMARFFETSMPPVYAIIVANCRSWTEFYLLISODPVEEFVEFGFESKVGVLSPALPH 348
07 244 IVAWIGVWGVVLADDELLTKKRELLT-VRKATITLGSPLSALVSLPYLNSGYITATL 302
Db 349 LVMTIIVPIGGIADPLFKSRIMSTTNTRKMLKNCGGGMEATLLLVGYSHSGV-AISF 407
07 303 LTLSCGLSTLCSGSIYIVNDIAPRYSFSLGASGRCSIIAPVIVPVSGFLLSQDEFG 362
Db 408 LVLANGFSGFPAISGNNVHNDIARVYASILMGISNGVGTLSGMCPILVGMATKHKTRRE 467
07 363 WRNVEFLLEPAVNLGLLEFYLLIGEADVQEMAEKRLT 399
Db 468 MQYVELLIASLVHGYGVIFYGFASGEKQPMAPPEMS 504

RESULT 5
W70500
ID W70500 standard; Protein; 560 AA.
AC W70500:
DE 18-JAN-1999 (first entry)
DE Human sodium-lithium countertransporter BNPI.
KM Sodium-lithium countertransporter; sodium-phosphate cotransporter;
KM BNPI; human; lithium therapy; manic depression.
OS Homo sapiens.
PN M09838203-A1.
PD 03-SEP-1998.
PF 11-FEB-1998; U02875.
PF 27-FEB-1997; U5-039462.
PA (UYEM-) UNIV EMORY.
PI Gunn RB, Timmer RT;
DR WPI; 98-520759/44.
DR N-PSDB; V33503.
PT New isolated lithium-sodium counter-transporter DNA - used to
PT develop products for evaluating lithium-sodium transport in
PT erythrocytes, particularly for lithium therapy in manic depression.
PS Claim 8; Page 45-47; 64pp; English.
CC This polypeptide comprises the brain-specific human lithium-sodium
CC countertransporter (LSTC) BNPI. A DNA sequence for human BNPI is
CC provided (see V33503). LSTCs such as BNPI provide the
CC physiological mechanism for the extrusion of lithium from cells,
CC i.e. it regulates the cell concentration of lithium. Its activity
CC determines the therapeutic effect of lithium. The invention
CC provides a simple molecular biological test for the ability of
CC cells to extrude lithium. The LSTCs have significance for
CC determining the responsiveness of humans with mental disorders,
CC including manic depressives, to treatment with lithium salts.
CC Probes and primers for BNPI, P1R-1 (see W70498), P1R-2 (see
CC W70499) can be used in diagnostic tests useful for genetic
CC screenings to predict whether a patient will respond to lithium
CC treatment. The test is also a screen for susceptibility to, and
CC extent of, manic depressive illness, and is suitable for screening
CC newborns.
SQ Sequence 560 AA;

```

Query Match	23 18	Score 479;	DB 1;	Length 560;
Best Local Similarity	28.28	Pred. No. 1.2e-42;		
Matches 112: Conservative	68;	Mismatches 155;	Indels 62;	Gaps 7;
QY	12	VPSLSARGIALVLFHFCNFTTIAQVNIINNTMVAWVNSTSPQSOLNDSSEVLVYDSFEG	71	
DB	161	IPSAARVHVGCIY-----FVRILGILVGCATGYPAC-----	191	

QY 72 LSKAPKSLPANSSILGGFPAIERNWNGPPOERSRLCSIALSGMLGCTFAILLGFISETL 131
 DB 191 -----HGIISSKMAPPLERSRLATFAFCGSAGAVVAMPPLAGLVYCS 232
 QY 132 GMPFEYFIYFGVGCYCCLLMFVYIYDDVSYVWISTSEKEYIISLKKOQVSSKQ---PL 188
 DB 233 GMSVFYIYGSGLFWYLFVWLLVSYESPALHPSISEEKRYI---EALIGRSALIMPL 288
 QY 189 -----PIKAMLRSLPIMSICLCFSHOWLVSTWVYIPTYISSYHVNRDNGLSALPE 243
 DB 289 TKFSIPMRFRFTSMVYALIVANFCRSWTFYLLISQPDYEEVGFELSKYGLVSLAPH 348
 QY 244 IVAMVWGVGYLADFLTKFRRLT-VRKATITLGSPLSSALIVSLPYNSGYITATL 302
 DB 349 LMTIIVIGQIADFLRSRIMSTTVNRKLMNCGFGMEATLLVGYSHSKGV-AISF 407
 QY 303 LTLSCGLTLOSGIYIVNLDIAPRYSFIMKASRGFSSIAPIVPTVSGFLISDPERG 362
 DB 408 LVLAVGFGFALSGFNVNHLDIAPRYASILMGISNGVGTLSGMWCPITIVGAMTKHKTREE 467
 QY 363 WRNVFFLLFAVNLGLLFYILFEGADVQEMAKERKIT 399
 DB 468 WQYVFLIASLVHYGVITFYGVFASGEKQPMAPPEMS 504

RESULT 6
 ID W88523 standard; protein; 576 AA.
 AC W88523:
 DT 05-MAR-1999 (first entry)
 DE Eat-4 protein amino acid sequence.
 KW Eat-4 protein; CAMP regulatory gene; neurodegeneration; CAMP;
 KW cyclic adenosine monophosphate; reporter gene; Alzheimer's; Parkinson's;
 KW Huntington's disease; amyotrophic lateral sclerosis; multiple sclerosis;
 KW stroke; epilepsy; GTP-binding protein G alpha s; Gas;
 KW adenyllyl cyclase gene; Gas-induced cell death; CAMP level.
 OS Caenorhabditis elegans.
 PN W09853856-A1.
 PD 03-DEC-1998.
 PF 29-MAY-1998; U11058.
 PR 29-MAY-1997; US-864785.
 PA (GENO) GEN HOSPITAL CORP.
 PI Hart AC, Kaplan JM, Oppenheimer AJ;
 DR WPI: 99-045345/04.
 PT Identifying agents for treating or preventing neurodegeneration -
 PT by assessing their effect on CAMP regulatory genes or promoters.
 PT useful for, e.g. potentially treating Alzheimer's disease
 PS Disclosure; Fig 3; 66pp; English.
 CC The present sequence represents an eat-4 protein of *Caenorhabditis*
 CC elegans. The eat-4 gene is a CAMP regulatory gene. The specification
 CC describes the identification of compounds for treating (or preventing
 CC onset of) neurodegeneration. The method comprises treating a cell having
 CC a reporter gene linked to a CAMP (cyclic adenosine monophosphate)
 CC regulatory gene or promoter, with a test compound and measuring
 CC expression of the reporter gene. Such compounds are potentially useful
 CC for treating and preventing Alzheimer's, Parkinson's and Huntington's
 CC diseases, amyotrophic lateral sclerosis, multiple sclerosis, stroke or
 CC epilepsy. The methods are based on the observation that, in the nematode
 CC *Caenorhabditis elegans*, constitutive activation of the GTP-binding
 CC protein G alpha s (Gas) induces neurodegeneration, and that mutations in
 CC some genes, including adenyllyl cyclase-type genes, block Gas-induced cell
 CC death, suggesting that the effect of Gas is associated with changes in
 CC CAMP levels.
 SQ Sequence 576 AA;

Query Match 22.5%; Score 466.5; DB 1; Length 576;
 Best Local Similarity 32.3%; Pred. No. 2,7e-41;
 Matches 101; Conservative 65; Mismatches 134; Indels 13; Gaps 6;
 QY 92 IWERNGPPOERSRLCSIALSGMLGCTFAILLGFISETLMPFYVYIFGVGCYCCLLW 151
 DB 198 WMRVAPMERSKLATVATFTGSYAGAVGLPLSLFVSVSWMAAFYLYGVGYVIMILW 257

QY 152 FVYIYDDVSYVWISTSEKEYIISLKKOQV---SSKOP-----LPKAMLRSLPIMSICLG 205
 DB 258 FCVTFEKAPFHHTISOEKIFL-----EALIGHVSTHTPTINSIPKALVTSKPVAILVA 313
 QY 206 CFSHOWLVSTWVYIPTYISSYHVNRDNGLSALPEPIVAMVIGMWGYLADFLTKRF 265
 DB 314 NFARSMTFYLLQNLQTFYMKREALGKRIADSGLLAIPHLVNGCVVLMGQGLADVLRSMKI 373
 QY 266 RLIT-VRKATITLGSPLSSALIVSLPYNSGYITATLITLSCGLSTLCOSGIYIVNLDI 324
 DB 374 LSTTVARKIFMCGGFGGAAFMILVAYTTSD-TTAIMLILAAVSGFALSGFNVNHLDI 432
 QY 325 APRYSFIMKASRGFSSIAPIVPTVSGFLISQDEFGMRNVFFLLFAVNLGLFYLI 384
 DB 433 APRYAILMGFSNGIGITAGLTCPTVTAFTAHS-KHGMTSVFLASLHFTGTFTAVY 491
 QY 385 GEADVQEMAKERK 397
 DB 492 ASGELQEWAEPRE 504

RESULT 7
 ID W14439 standard; protein; 482 AA.
 AC W14439:
 DT 15-MAY-1997 (first entry)
 DE Protein involved in cephalosporin C biosynthesis.
 KW cephalosporin C; biosynthesis; fermentation.
 OS *Acromonium chrysogenum*.
 PN J09009966-A.
 PD 14-JAN-1997.
 PF 03-JUL-1995; 167461.
 PR 03-JUL-1995; JP-167461.
 PA (ASAH) ASAH KASEI KOGYO KK.
 DR WPI: 97-126424/12.
 DR N-PSDB: T62652-53.
 PT *Acromonium chrysogenum* gene involved in biosynthesis of
 PT cephalosporin C - used to improve fermentation ability of A.
 PT *chrysogenum*
 PS Claim 2; Page 9-11; 21pp; Japanese.
 CC This protein, encoded by DNA derived from *Acromonium chrysogenum*, is
 CC involved in biosynthesis of cephalosporin C. The gene involved in
 CC biosynthesis of cephalosporin C or its cDNA can be used to improve the
 CC fermentation ability of *Acromonium chrysogenum*.
 SQ Sequence 482 AA;

Query Match 8.2%; Score 169.5; DB 1; Length 482;
 Best Local Similarity 24.2%; Pred. No. 9.9e-10;
 Matches 81; Conservative 62; Mismatches 117; Indels 75; Gaps 17;
 QY 100 QERSRLCSIALSGMLGCTFAILLGFISETLGMPEFYIFGVGCYCCLLMFVYIYDDP 159
 DB 165 EERGAAMAIWTFEGLMGPVAVGAYLAELAKGMWVWVAIIGGFTTGAFILARE-- 223
 QY 160 VSYPIISSEKEYIISLKKOQVSSKQPL-----PIKAMLRSLP 198
 DB 223 -TTPVLLQKR---VNRLOQETGN---PLTSALADISSRRARISRSYRRLVLLFRS-P 274
 QY 199 I---WSICIG-CFSHOWLVSTWVYIPTYISSYHVNRDNGLSALPEPIVAMVIG- 253
 DB 275 IYVLFESVFIIVAVFSYQFL---LFTYIPVFEIYDFLSGQIG-LSYLGIAAGLLGNRIF 330
 QY 254 GYLADFLTKR-----RLITVRKIATITLGSPLSSALIVSLPYNSGYIT----- 299
 DB 331 GOASDRILSKSKSGMEKLEPERLPLM-----IP-GARCIPMCFEYIYGMATYKLMW 380
 QY 299 -----ATALLTLSCGLSTLCOSGIYIVNLDIAPRYSFIMKASRGFSSIAPIVPTVSGF 353
 DB 381 MPMICATSLIGLIGNLISMT--IQVYIVDYTYTLYSASALAAATLRSDFGAFILP-LAGP 436
 QY 354 LLSQDEPFGMRNVFFLLFAVNLGLFYLI-FGEA 387

Db 437 PLYDALGLMGNSTLGFIAVALIPVPLEFRYGEA 471

RESULT 8

W97820

AC W97820 standard; Protein: 462 AA.

DE 21-MAY-1999 (first entry)

KW Staphylococcus aureus tetracycline resistance protein.

KW Tetracycline resistance; antibiotic; antibacterial; screening;

KW infection; otitis media; endocarditis; abscess; conjunctivitis;

KW epididymitis; toxic shock syndrome; sepsis; septic arthritis;

KW diagnosis; therapy; vaccine.

OS Staphylococcus aureus.

PN EP-893499-A2.

PD 27-JAN-1999.

PR 10-JUL-1998; 305495.

PR 23-JUL-1997; US-898976.

PA (SMIK) SMITHKLINE BEECHAM CORP.

PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Burnham MKR, Lomello MA, Warren PV;

PI WPI: 99-097783/09.

DR N-PSDB: X07357.

PT New Staphylococcus aureus tetracycline resistant protein and gene -

PT useful as diagnostic reagents and for prevention and treatment of

PT staphylococcal infections, including toxic shock syndrome and otitis

PT media

PS Claim 12: Page 27-29; 35pp; English.

CC This is the amino acid of a novel tetracycline resistance protein

CC (TR) of Staphylococcus aureus WCUH29 (NCIMB 40771) that is involved

CC in sugar or drug transport across bacterial cytoplasmic membranes.

CC TR polynucleotide (see X07357) has been isolated from an S. aureus

CC DNA library. Vectors comprising the claimed polynucleotide, host

CC cells and a process for producing the TR protein are claimed. TR

CC polypeptides and polynucleotides are useful for diagnosing diseases

CC related to over or under expression of TR protein by identifying

CC mutations in the TR gene, or determining TR polypeptide or mRNA

CC expression levels due to an infection of an organism with the TR

CC gene (claimed). They can diagnose the stage and type of infection.

CC TR is also useful for screening for compounds which affect the TR

CC activity. These can be used in treatment to inhibit (antagonist

CC 1.e. antibacterial drugs) or enhance (agonist) TR activity. Direct

CC administration of TR polypeptides can be used to treat conditions

CC associated with a lack of TR (claimed). Direct administration of

CC antisense oligonucleotides prevents expression. TR polypeptides

CC (administered directly, in a vector (gene therapy) and as a vaccine)

CC and antibodies induce an immune response to immunise and prevent

CC disease (claimed). Diseases diagnosed, prevented or treated

CC include: bacterial, especially S. aureus, infections of the

CC respiratory tract (e.g. otitis media), cardiac (e.g. infective

CC endocarditis), gastrointestinal (e.g. splenic abscess), CNS (e.g.

CC cerebral abscess), eye (e.g. conjunctivitis), kidney and urinary

CC tract (e.g. epididymitis and toxic shock syndrome), skin (e.g. TR

CC wound infection), and bone and joint (e.g. septic arthritis). TR

CC polypeptides, polynucleotides and their (ant)agonists can prevent

CC adhesion of bacteria to matrix proteins, and are useful for use on

CC wounds and body implants to prevent bacterial infection. TR is

CC transcribed during both chronic and acute infections.

SQ Sequence 462 AA;

Query Match 5.9%; Score 122.5; DB 1; Length 462;

Best Local Similarity 17.7%; Pred. No. 9.6e-05;

Matches 76; Conservative 86; Mismatches 158; Indels 109; Gaps 18;

QY 20 YGIALVHRCNFTTIANQYIMN---ITWAMVNSTPSQSLNDSSVLPVDSNGSKAP 76

DB 92 YVILIT---FGSIIGNFQHSPLTLVGRITQTA-----GIAAAE 129

QY 77 KSLPAKSSILGQFAIM-ERMGPPOERSRLCSIALSGMLGCFATILGFISETLGM-- 134

DB 130 T-----LYVIYVAKYLSKEDQKTYLGLSTSSYSLSLVIGTILSGFIITYLHMWN 178

QY 134 -----PEVEYI-----FGVGCVCCLMFVYVDDPVSPMISNS 168

DB 179 MFLIALIVFTLPFLFKLPKRNNTKKAHLDFVGLVATITATWMLF--ITNFMNL---- 234

QY 169 EKEYIISL-----KQVGSCKOPLPIKAMLRSLPWSICLGCFSHOWLVSTWVYIPT 222

DB 234 ---YMGALIAITVFAIYIKNAQRPLVNSKSPQNKRYASFLLVFMVAIQGLYTFPF 290

QY 223 YISSYHVHNRDGLSALPFIYAVYGVGLADFLTKFRLLITVKRIATILGSLPS 282

DB 291 IMEQITHLDPTSLILPVGYIAVYVAGALSGKIGELNSKQAITIILIALSL-ILPA 349

QY 283 -----SALIYSLPYLNSGYITATALLTSLGSLTLCOSGIYINVLAPRSPMLKA 335

DB 350 FAVGNHISIFVISMIFPAGSFALMTAPL-LNEAIKT-----IDLNTGVAIGFYNLINIV 403

QY 336 SRGFSIAPIVPTVSGFLSODP-----EFGWNRVFFLLFAVNLGLFLYLGEND 388

DB 404 A---VSVGIAIAAALIDFRAINPFGNDALSSHFG--IILILGLMSIVGLVLFYI----- 454

QY 389 VOEMAKERK 397

DB 454 LNRWTOSEK 462

RESULT 9

W56446

AC W56446 standard; Protein: 2969 AA.

DE 31-JUL-1998 (first entry)

DE Fragment HCU2141 of a new Hepatitis virus g protein.

KW Hepatitis g virus gene; diagnosis; treatment; Hepatitis g virus disease.

OS Hepatitis g virus.

FN Key Location/Qualifiers

FT MISC_difference 1388..1407

FT "nucleotides encoding these residues not

FT given in the specification"

FT MISC_difference 1448..1467

FT "nucleotides encoding these residues not

FT given in the specification"

FT "nucleotides encoding these residues not

FT given in the specification"

FT J10108685-A.

PD 28-APR-1998.

PR 10-AUG-1997; 227387.

PR 10-AUG-1996; JP-227639.

PA (BMLB-) BML KR.

DR WPI: 98-304974/27.

DR N-PSDB: V23080.

PT New hepatitis G virus gene - useful for diagnosing and treating

PT diseases caused by virus

PS Claim 1: Pages 108-112; 128pp; Japanese.

CC The present sequence represents a fragment of a new Hepatitis g virus

CC protein (see also W56441-50 for other fragments). RNA was synthesised

CC from the serum of nine patients judged positive for Hepatitis g virus

CC and cDNA synthesised from this RNA. The cDNA was used as a template in

CC several PCR reactions to isolate fragments of the new gene. The gene

CC may be useful for diagnosing and developing treatments for Hepatitis g

CC virus diseases.

SQ Sequence 2969 AA;

Query Match 5.4%; Score 112.5; DB 1; Length 2969;

Best Local Similarity 25.0%; Pred. No. 0.017;

Matches 97; Conservative 35; Mismatches 133; Indels 123; Gaps 23;

QY 79 LPAKSSILG-GQFAIMERMGPPOERSRLCSIALSGMLGCFATILGFISETL-----L 131

DB 637 LPEVSMILGLANLVYFRWMPQ---RLMFLVLMKLARKGAFPLALMG-ISAIRGRSVL 692

QY 132 GMPFVF-----YIFGVGCVCCLMFVYVDDPVSPW-----ISTSEKEYIISL 177

DB 693 GAEFCFDVTFVEVDTSVLGNV-VASVVAAMIALLSMSAGCMRKAVIYRWCKGY--QAL 749

OY	178	KQOV-----	GSSKOPLPKAMRLSPITS-----	ICLCGF-SHOWLVSTMYV	218
Db	750	KORVAFSELGEGREPTKPLTIAMCLASY-IPMDAVMLVVGVLVEGLFDALDALEBLELV			808
OY	219	YIPFYISSEYHVHNRDNGLLSALPEIY-----	AMVGMV-----	GGYIAD-----	F 259
Db	809	SRPS-----	LRLRLRVCECCVMGEKAKTITRIIVSKMCARAYILFDHMGSF		853
OY	260	LITKFRLIYRKATIATIIIGSLPSSALIYSEPLYNSG-YITATATLITISGLSLSTLCOSGIY			318
Db	854	SRAVKERL-----	LEMDAALVPLSEFTPTDCRIIRDAARTLSCG-----	QCWVGLP	898
OY	319	INVLIDIAPIRSSFLMGASRCGFSSIAPIVPIV-----	SGF-----	LLSODPEFGW	363
Db	899	V-----	VARRGVEVLIGVQDVNHLPPGCVIPIAPVYIRCGGGLGYTKAKLALTGRDPLHP		954
OY	364	RNVFFLEFAVN-----	LLGLLEFLYLIIG		385
Db	955	GNVAVLGTATSRSMGTCINGILLFTFTFHG			982

RESULT 10
ID W56450 standard; Protein; 2965 AA.
AC W56450;
DT 31-JUL-1998 (first entry)
DE Fragment F1029 of a new Hepatitis virus g protein.
KW Hepatitis g virus gene; diagnosis; treatment; Hepatitis g virus disease.
OS Hepatitis g virus.
PN J10108685-A.
PD 28-APR-1998.
PF 10-AUG-1997; 227387.
PR 10-AUG-1996; JP-227639.
PA (BMLB-) BML KK.
DR WPI: 98-304974/27.
DR N-PSDB: V20384.
PT New hepatitis G virus gene - useful for diagnosing and treating diseases caused by virus
PS Claim 1; Pages 123-127; 128pp; Japanese.
CC The present sequence represents a fragment of a new Hepatitis g virus protein (see also W56441-50 for other fragments). RNA was synthesised from the serum of nine patients judged positive for Hepatitis g virus and cDNA synthesised from this RNA. The cDNA was used as a template in several PCR reactions to isolate fragments of the new gene. The gene may be useful for diagnosing and developing treatments for Hepatitis g virus diseases.
SQ Sequence 2965 AA;

	Query Match	5.4%	Score 111.5:	DB 1;	length 2965;
	Best Local Similarity	23.2%;	Pred.	No. 0.022;	
	Matches 91; Conservative		38; Mismatches	132;	Indels 131; Gaps 22.
OY	LPAKSSILG-GQFAIWERNGPPOERSRLCSIALSGMLGCCFPAIIIGFISET-----L	131			
Dd	637 LPEVSMITGLAMLVIFERFMGPQ---RIIMFLVIWMKLARCAFPALILMG-ISAIRGRSTVL	692			
OY	132 GMPPVF-----YIFGVGCVCCLLMNVVIVYDDPVSYIPW-----ISTSEKEYIIISL	177			
Dd	693 GAEPCEVDVTFFEVDTSVLGNV-VASVLAAMIALLISSAGCGMKRAKAVITYWCQGY-QAI	749			
OY	178 KOOV-----GSKKPLDIKAMLNSLPITSICLGCFSHQMTVSTNNVVYIPTYYISSVHVH	230			
Dd	750 RORVASPLGEGRPKTPILT-----AM-CIA-SYIWEDAVNLVAVALILE----	794			
OY	231 NTRDGLSLAPPIVMAYGMVGVLADLFLLTKKF-RLIITYRKATIIIGLSPLSALIYS	288			
Dd	794 -----GLEPALD--WA-----LEELIVSRPSLRKLARIYVECCVMAGEKATTIRIAYS	837			
OY	289 LPYLNSGY-----ITATALTLSCGSTLCO	314			
Dd	838 KMCARGAAYFDHMGSFSRVRKERLEEMDALPLESTRTDCRIIRPDAAFTLSCG--QCQV	894			

QY 315 SCGIYINVLDIRKSSFLPMGASRGESSIAPIVTV-----SGP-----LLSDP 359

Db 895 MGLP-----VARRGDEVILGIVQDVNHHLPPEFVAPYAVIRCKGKPLGVTKAALIGRDP 950

QY 360 EFGWENVEFLLFAVN-----LLGLLFYLLFG 385

Db 951 DLHPGNVAVLGTATSRSMGTCLINLGLLTTFNG 982

```

RESULT 11
ID W56445
AC W56445;
DT 31-JUL-1998 (first entry)
DE Fragment HGJ1916 of a new Hepatitis virus g protein.
KW Hepatitis g virus gene; diagnosis; treatment; Hepatitis g virus disease
OS Hepatitis g virus.
FT key Location/Qualifiers
FT Misc_difference 2518
FT /note= "not specified; encoded by CCG"
FT Misc_difference 2824
FT /note= "Phe encoded by CTC"
FT Misc_difference 2857
FT /note= "Arg encoded by TGC"
FT Misc_difference 2865
FT /note= "nucleotides encoding this residue not given"
PN J10108685-A.
PD 28-APR-1998.
PE 10-AUG-1997; 227387.
PR 10-AUG-1996; JP-227639.
PA (BMRB-) BML KR.
PA WP1: 98-304974/27.
DR N-PSDB; V23079.
PT New hepatitis g virus gene - useful for diagnosing and treating
PT diseases caused by virus
PS Claim 1; Pages 104-108; 128pp; Japanese.
CC The present sequence represents a fragment of a new Hepatitis g virus
CC protein (see also W56441-50 for other fragments). RNA was synthesised
CC from the serum of nine patients judged positive for Hepatitis g virus
CC and cDNA synthesised from this RNA. The cDNA was used as a template in
CC several PCR reactions to isolate fragments of the new gene. The gene
CC may be useful for diagnosing and developing treatments for Hepatitis g
CC virus diseases.
SQ Sequence 2970 AA;

```

```

Query Match      5.4%; Score 11.5; DB 1; Length 2970;
Best Local Similarity 23.0%; Pred. No. 0.02;
Matches 90; Conservative 38; Mismatches 133; Indels 131; Gaps 21;

Qy 79 LPAKSIIG-GGFALMERKGPPOERSRSLNLSGMLGCFNALIGFISPT-----L 131
   || ||||| :||| || ||| :| :| :| :| :| :| :| :| :| :| :| :|
Db 637 LPPVSMITGLANLVYFRMGPO--RLMFLVLMKIAKARAPFLALLMG-ISTRORTSYL 692
   ||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 132 GMPFVF-----YIFGCGVCVCCLLMNVVYIDDPVSYPW-----ISTSEKXYIISL 177
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 693 GAEFCEDVFEYEDTVSLGVN-VASVYAMAILSSMSAGCMHKAVIYRTWCkgV--QAL 749
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 178 KOQV-----GSSKQPLDIKAMLSLPIWISICLGCSHQWLVSTMVVYIPIYISSVHY 230
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 750 RQVYVSPLEGRSPAKPLTI-----AM-CLA-sYIWPDAWMLVVGVLVLF----- 794
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 231 NIRDNGLLALFFIVAWVIGWVGYLADPLRLKRF--RLITPRKLIATIGSLPSALIVS 288
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 794 ----GLEFDALD--WA-----LEILLVSRPSLIRRAVVECCWMAGEKATVRLVS 837
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 289 LPIYNSG-----ITATYLLTSCGSLTLCQ 314
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 838 KMKARAVILEDHGGSRAVKERLEMDALALPLSTFRTRDCLINDAANTLSCG---QCV 894
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Qy 315 SGIIYNVLDIAPRYSSEFLMGASRGSSIAPIVPT-----VSGFLISDP 359
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
Db 895 MGIPV----VARGGDEVILGLQDVNHLPRGPGVPTPAVIRMGCGSGSLGVTAAALTRGP 950
   :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```


OY	132	GWFVF-----YINGVGVCCCLMFVIYDDPSYPW-----ISTSEKEYIISSL	177
Db	809	GAEECFDVEFVDTSVLGW-VASVAMALALLSSMAGCKKRAIITRWCKGY--QXL	865
OY	178	KOOV-----GSKODLPFKAMLRSLPLPISICIGCSHOWLSTVMVYIPTVISSYHV	230
Db	866	RQRYVNSPLEEGPRPKPLTI-----AW-CLA--SYMPPAVMLVVAMVLE----	910
OY	231	NINDGLLSLPTIVMVGWGGYLADFLITKKF-RLTVRKIATILCSLPSSALIYS	288
Db	910	-----GLEFDALD---WA-----LEEILLVSRLRIARAVECCWAGEKAITYRLVS	953
OY	289	LPIYNSGY-----TFAFLLTSCGLSTLCQ	314
Db	954	KMCARAVILEDHMGFSRBAVKERLEWDALAEKLSTRTCDRIIRDAARTLSCG---QC	1010
OY	315	SGIIVNLDIAPRYSFSLNGASR-----GFSSIAPVIVPYT-SGP-----LISDP	359
Db	1011	MGIIPV-----VARRGDEVILIGVFDVNHLPGFXPTAPVIVIRCGKGPLGTVAALNGRDP	1066
OY	360	EFGWRVFFLLPAVN-----LLGLLFYLIIFG	385
Db	1067	DHPGNMVVLGTATSRSMGTCINLGLETTFEFG	1098
RESULT	15		
R99353	ID	R99353 standard; Protein: 470 AA.	
AC	R99353;		
DT	16-DEC-1996	(first entry)	
DE	Human amine transporter.		
KW	Amine transporter; neurotransmitter; neurological disease;		
KW	Alzheimer disease; Parkinson disease; therapy; diagnosis;		
OS	Homo sapiens.		
FH	Key	Location/Qualifiers	
FT	misc_difference 402		
FT	/note= "unidentified amino acid"		
PN	WO9627009-A1.		
PD	06-SEP-1996.		
PF	01-MAR-1995; UO2645.		
PR	01-MAR-1995; WO-UO2645.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Cao L., Li Y., Rosen CA;		
DR	WPI: 96-412775/41.		
DR	N-PSDB: Tj3417.		
PT	Human amine transporter (HAT), stimulates uptake of		
PT	neurotransmitter(s) - useful for treating and diagnosing		
PT	neurological disease, e.g. Parkinson's or Alzheimer's disease.		
PS	Claim 1: Page 39-41; 58pp; English.		
CC	Human amine transporter (R99353) stimulates uptake of		
CC	neurotransmitters. Its amino acid sequence was deduced from		
CC	a cDNA clone (Tj3417) isolated from CDNA library derived from		
CC	a human adrenal gland tumour. Recombinant amine transporter can		
CC	be produced in prokaryotic or eukaryotic host cells, e.g. E. coli,		
CC	COS and Spodoptera frugiperda Sf9 insect cells. It may also be		
CC	used to screen for (ant)agonists of amine transporter activity.		
CC	Agonists can be used to treat disorders, such as Parkinson's and		
CC	Alzheimer's diseases, related to under-expression of amine		
CC	transporter activity. Antagonists (including soluble amine		
CC	transporter and antibodies) can be used to treat conditions related		
CC	to excess amine transport, such as epilepsy, depression,		
CC	schizophrenia.		

Query Match	5.2%	Score 107.5;	DB 1;	Length 470;
Best Local Similarity	24.4%	Pred. No. 0.0039;		
Matches 64;	Conservative	37;	Mismatches 100;	Indels 61;
				Gaps 11;
101	ERSRCSIALSGMLIGCTATILIGGFISETIG--WPFVFIIF-----GGVGCACCLINREVV	154		

```

OY 155 IYDDPVSYPMISTSEKEYIISLKOQVSS-KOPLPIKAMLRSPISICIGCFSHOWLV 213
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 284 SPESAKGTPLFMLKDPYILVA-----AGSSCFANMGVAIIPLPTIMQTMG-SPKWOL 338
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 214 STWVVIPTIYISSVYHVNIRDNGILSALPPIVAMVIGM-VGYLADFLTKKFRLLTVRK 272
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 G-----LAFLPASVSYLIGTNLFVLANKM-----GRW 366
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 273 IATILGSL-PSSALIVSLPYLNSGYITATALLTSLGSLTLCQSGIYINVLDIAPRYSS 330
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 367 LCSLIGMLVGTSLCLCVLAHKNFGLIGPNAGLGLXIGWESSMMPIMSHLVD--PRHTS 424
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 331 -----FLMGASRGFS 340
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 425 VYGSVHAIAADVAFCMGFATIGYS 446
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: May 25, 2000, 14:11:15
 Job time: 2352 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 13:38:17 ; Search time 40.12 Seconds

(without alignments)
144.302 Million cell updates/sec

Title: US-09-391-958-1

Sequence: 1 MOWDETLPKRVPSLCSARY.....LIFGEADVQEMAKRKRLRL 401

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: /cgn2_6/pdata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/pdata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/pdata/1/1aa/6.COMB.pep:*
4: /cgn2_6/pdata/1/1aa/PCITUS.COMB.pep:*
5: /cgn2_6/pdata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2076	100.0	401	2	US-08-805-118-1	Sequence 1, App1
2	1991.5	95.9	480	2	US-08-724-394A-11	Sequence 11, App1
3	964	46.4	467	2	US-08-805-118-3	Sequence 3, App1
4	957.5	46.1	480	2	US-08-724-394A-9	Sequence 9, App1
5	722.5	34.8	470	2	US-08-724-394A-10	Sequence 10, App1
6	479	23.1	560	1	US-08-647-484-2	Sequence 2, App1
7	479	23.1	560	1	US-08-647-481-2	Sequence 2, App1
8	479	23.1	560	1	US-08-430-033A-2	Sequence 2, App1
9	479	23.1	560	4	PCT-US96-05792-2	Sequence 2, App1
10	475	22.9	560	2	US-08-805-118-4	Sequence 4, App1
11	122.5	5.9	462	2	US-08-898-976-2	Sequence 2, App1
12	122.5	5.9	462	2	US-08-898-976-4	Sequence 4, App1
13	106.5	5.1	2873	1	US-08-466-033-15	Sequence 15, App1
14	106.5	5.1	2873	1	US-08-638-911A-2	Sequence 2, App1
15	106.5	5.1	2873	2	US-08-444-733-15	Sequence 15, App1
16	106.5	5.1	2873	2	US-08-464-134-15	Sequence 15, App1
17	106.5	5.1	2873	2	US-08-461-361-15	Sequence 15, App1
18	106.5	5.1	2873	2	US-08-485-910-15	Sequence 15, App1
19	106.5	5.1	2873	4	PCT-US95-06266-15	Sequence 15, App1
20	105	5.1	426	5	5268463-8	Patent No. 5268463
21	104.5	5.0	470	1	US-08-471-496-2	Sequence 2, App1
22	104.5	5.0	470	1	US-08-894-840-2	Sequence 2, App1
23	104.5	5.0	2910	1	US-08-466-033-183	Sequence 183, App
24	104.5	5.0	2910	2	US-08-444-733-183	Sequence 183, App
25	104.5	5.0	2910	2	US-08-464-134-183	Sequence 183, App
26	104.5	5.0	2910	2	US-08-461-361-183	Sequence 183, App
27	104.5	5.0	2910	2	US-08-485-910-183	Sequence 183, App
28	104.5	5.0	2910	4	PCT-US95-06266-157	Sequence 157, App
29	103.5	5.0	488	2	US-08-928-692-10	Sequence 10, App1

30	103.5	5.0	514	1	US-08-063-552-13	Sequence 13, App1
31	103.5	5.0	514	4	PCT-US93-05704-13	Sequence 13, App1
32	101.5	4.9	537	2	US-08-647-397-2	Sequence 2, App1
33	101	4.9	456	5	5432081-7	Patent No. 5432081
34	98	4.7	457	2	US-08-882-704A-6	Sequence 6, App1
35	97.5	4.7	457	5	5268463-7	Patent No. 5268463
36	96	4.6	515	1	US-08-063-552-4	Sequence 4, App1
37	96	4.6	515	4	PCT-US93-05704-4	Sequence 4, App1
38	94.5	4.6	521	1	US-08-063-552-2	Sequence 2, App1
39	94.5	4.6	521	4	PCT-US93-05704-2	Sequence 2, App1
40	94	4.5	3287	2	US-08-477-451-7	Sequence 7, App1
41	93	4.5	595	2	US-08-677-049-11	Sequence 11, App1
42	92.5	4.5	465	1	US-08-471-496-9	Sequence 9, App1
43	92.5	4.5	465	2	US-08-894-840-9	Sequence 9, App1
44	92.5	4.5	580	2	US-08-677-049-12	Sequence 12, App1
45	92	4.4	397	1	US-08-098-141-2	Sequence 2, App1

ALIGNMENTS

RESULT 1
US-08-805-118-1
; Sequence 1, Application US/08805118
; Patent No. 5985604
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0221 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 401 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: BRAINT02
CLONE: 754412
US-08-805-118-1

Query Match 100.0%; Score 2076; DB 2; Length 401;
Best Local Similarity 100.0%; Pred. No. 9.5e-212;
Matches 401; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MOVDETLIPRKVPISCSARYGIALVLFHCNFTTIAQNVIMNTVMVAVNSTSPQSOLNDS 60
D 1 MOVDETLIPRKVPISCSARYGIALVLFHCNFTTIAQNVIMNTVMVAVNSTSPQSOLNDS 60
QY 61 SEVLVDSRGSGSKAPKSLPAKSSILGGFALWERNKGPQERSRLCSIALSGMLGCFPA 120
D 61 SEVLVDSRGSGSKAPKSLPAKSSILGGFALWERNKGPQERSRLCSIALSGMLGCFPA 120
QY 121 ILIGFISSETLGMPEFYIFGVCVCCCLMFVYIYDDPVSPWISTSEKEIYISLKQ 180
D 121 ILIGFISSETLGMPEFYIFGVCVCCCLMFVYIYDDPVSPWISTSEKEIYISLKQ 180
QY 181 VSSSQOPLPIKMLNSLPISICLCFSGHQLVSTMVVIIPYISSVYHVNIRNGLLSA 240
D 181 VSSSQOPLPIKMLNSLPISICLCFSGHQLVSTMVVIIPYISSVYHVNIRNGLLSA 240
QY 241 LPFIYAWYIGMWGYLADPLTKKRLITVRKIATILGSLPSALIVSLPIYNSGYTAT 300
D 241 LPFIYAWYIGMWGYLADPLTKKRLITVRKIATILGSLPSALIVSLPIYNSGYTAT 300
QY 301 ALLTSLGSLTLCQSGIYINVDIAPRYSSFLMGASRGSSIAPIYIPVSGFLLSODE 360
D 301 ALLTSLGSLTLCQSGIYINVDIAPRYSSFLMGASRGSSIAPIYIPVSGFLLSODE 360
QY 361 FGMNVFFLLFAVNLGLFLYIFGEADVQEWAKERKRL 401
D 361 FGMNVFFLLFAVNLGLFLYIFGEADVQEWAKERKRL 401

```

RESULT 2
US-08-724-394A-11
Sequence 11, Application US/08724394A
Patent No. 5872237

GENERAL INFORMATION:
APPLICANT: Feder, John N.
APPLICANT: Kromal, Gregory S.
APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereof
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND AND TOWNSEND AND CREM LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 480 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant

MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Region
LOCATION: 1..480
OTHER INFORMATION: /note= "NP74"
US-08-724-394A-11

Query Match 95.9%; Score 1991.5; DB 2; Length 480;
Best Local Similarity 82.9%; Pred. No. 1.1e-202;
Matches 398; Conservative 1; Mismatches 2; Indels 79; Gaps 3;

```

QY 1 MOVDETLIPRKVPISCSARYGIALVLFHCNFTTIAQNVIMNTVMVAVNSTSPQSOLNDS 60
D 1 MOVDETLIPRKVPISCSARYGIALVLFHCNFTTIAQNVIMNTVMVAVNSTSPQSOLNDS 60
QY 61 SE-----VLVDSFGGLSKA 75
D 61 SEXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXVLPVDSFGGLSKA 120
QY 76 PKSLP-----AKSSILGGFALWERNKGPQER 102
D 76 PKSLPXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXVLPVDSFGGLSKA 120
QY 121 PRSLPXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXVLPVDSFGGLSKA 120
D 121 PRSLPXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXVLPVDSFGGLSKA 120
QY 163 PWISTSEKEIYISLKQVSSKOPLPKAMRLSLPISICLCFSGHQLVSTMVVIIP 222
D 163 PWISTSEKEIYISLKQVSSKOPLPKAMRLSLPISICLCFSGHQLVSTMVVIIP 222
QY 241 PWISTSEKEIYISLKQVSSKOPLPKAMRLSLPISICLCFSGHQLVSTMVVIIP 300
D 241 PWISTSEKEIYISLKQVSSKOPLPKAMRLSLPISICLCFSGHQLVSTMVVIIP 300
QY 282 SSALIVSLPYINSGYITATALLTSCGLSTLCQSGIYINVDIAPRYSSFLMGASRGSS 341
D 282 SSALIVSLPYINSGYITATALLTSCGLSTLCQSGIYINVDIAPRYSSFLMGASRGSS 341
QY 361 SSALIVSLPYINSGYITATALLTSCGLSTLCQSGIYINVDIAPRYSSFLMGASRGSS 420
D 361 SSALIVSLPYINSGYITATALLTSCGLSTLCQSGIYINVDIAPRYSSFLMGASRGSS 420
QY 421 IAPVIVPVSGFLLSODEFGMRNVFFLLFAVNLGLFLYIFGEADVQEWAKERKRL 480
D 421 IAPVIVPVSGFLLSODEFGMRNVFFLLFAVNLGLFLYIFGEADVQEWAKERKRL 480

```

RESULT 3
US-08-805-118-3
Sequence 3, Application US/08805118
Patent No. 5985604

GENERAL INFORMATION:
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
TITLE OF INVENTION: PROSOPHATE CO-TRANSPORTER
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/805,118
FILING DATE: Filed Herewith
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

RESULT 5
US-08-724-394A-10
; Sequence 10, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Rudd, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereof
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fils, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 470 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Region
; LOCATION: 1..470
; OTHER INFORMATION: /note="NPT3"
; US-08-724-394A-10

Query Match 34.8%; Score 722.5; DB 2; Length 470;
Best Local Similarity 38.4%; Pred. No. 2.9e-68;
Matches 161; Conservative 48; Mismatches 95; Indels 115; Gaps 6;
QY 10 RVPSLCSARYGIALVHFCNFTTIAQVNIAMVNSTSPQSLNDSSEVLVPVDSF 69
DB 8 RKGDFCSLRYGLALIMHFSNFTITQVSLAIAMVNTTQOGLSNATDEGVADAF 67
QY 70 GGLSKAPKSLPAKSI-----LG----- 88
DB 68 NNSISIKIEFDTKASYVQWSPETOGIIFSSINYGIIILTPSGYLAGIFGAKMIGAIL 127
QY 88 -----GQFAIWERNGPPQERSRLCSTALS 111
DB 128 ISSILLETPLADPGVILVIMRTVQGMAGMAMTGOFTIMAWAPPLERSKLTITAGS 187
QY 112 GMLIGCTAILIGFISSETLAMPVFYIFGVCVCCILMFVVIYDDPVSYPIWISTSEKE 171
DB 188 GSAGSFILICVGLISQALSMPLFIYIFGSGTCVCCILMFVVIYDDPMHHCISVNEKE 247
QY 172 YIISLQOQVSSKQPLPIKAMLRSLPIWISICLCFSHOWLVSTMVVIPTIYISSVYHVN 231

DB 248 HTLSLQAQPPSSPGRAVPIKAMVTCPLPMAIFLGFSSHRLCTIITITPTIYSTLHVN 307
QY 232 INDNGILSLAPFTVAMVIGMVGGLADFLTRK-FRLTVRIATIL----- 278
DB 308 INDGVLSSLPFIAMASCTIILGQADFLSRNRLRLTLVRKLFSSLDMMQVSMWEXXXXX 367
QY 278 -----GLPSSALIVSLPY-LNSG-----YITFALLTLSCGLSTLC 313
DB 368 XXXXXXXXXXXXXSQGLDGSSQESXSLPLPLDSSSXXXXXXXXXXVRLSLVGMFSFC 426

RESULT 6
US-08-647-484-2
; Sequence 2, Application US/08647484
; Patent No. 5618677
; GENERAL INFORMATION:
; APPLICANT: NI, Binhui
; APPLICANT: Paul, Steven M.
; TITLE OF INVENTION: HUMAN BRAIN SODIUM SODIUM DEPENDENT INORGANIC
; TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,484
; FILING DATE: 14-MAY-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/430,033
; FILING DATE: 27-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 560 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-647-484-2

Query Match 23.1%; Score 479; DB 1; Length 560;
Best Local Similarity 28.2%; Pred. No. 2.3e-42;
Matches 112; Conservative 68; Mismatches 155; Indels 62; Gaps 7;
QY 12 VPSLCSARYGIALVHFCNFTTIAQVNIAMVNSTSPQSLNDSSEVLVPVDSFG 71
DB 161 IPSAARVHYGVY-----FVRIIGLVEGVYTAAC----- 191
QY 72 LSKAPKSLPAKSIILGQFAIWERNGPPQERSRLCSTALSGLMIGCTAILIGFISSETL 131
DB 191 -----HGIMSKAPPLERSRLATTAFCGSVAGAVAPPLAGVLYOYS 232
QY 132 GMPVFYIFGVCVCCILMFVVIYDDPVSYPIWISTSEKEIYISSLKQOQVSSKQ---PL 188
DB 233 GWSVFYVYGSFGIFWYLFMLVLYESBPALHPISISEERRYI---BDAIGESAKLMLNPL 288

```

Oy 189 ----PIKAMLSLPIWSSICLQCSHOMLVSTWVYIPIYISVYHVNIRDGLSALPE 243
Db 269 TKEFSFPMRRFFISMVYAIIVANFCRSKTFYLLISQDPYFEFVGFELISKGLVSALPH 348
Oy 244 IVAMVIGWGVGLADFLTKKRLIT-VKRIATILGLSPSSALIYSLPYLNSGITATL 302
Db 349 LVMITIVIPGGIADIFLBSRIRIMSTTNVRKLNCCGFGMEATLLLVGYSKSGV-AISF 407
Oy 303 LTLSCGLSTLCOSGIYINVLDIAPRYSFLMGASRGESSIADYIVPTVSGFLISQDPFG 362
Db 408 LVLAFGSGFALSGFNVNHLDIAPRYASIIAMISNGVGLSGWCPYIIGAMTKHKTREE 467
Oy 363 WRNVFELFAVNLGLLEFLFEGEADVOQMAERKRT 399
Db 468 WOYFELIASLVHYGVIFGVFASGEKOPMAPEEMS 504

RESULT 7
Oy US-08-647-481-2
Db Sequence 2, Application US/08647481
Patent No. 5618918
GENERAL INFORMATION:
APPLICANT: NI, Binhui
APPLICANT: Paul, Steven M.
TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/647,481
FILING DATE: 14-MAY-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/430,033
FILING DATE: 27-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-100006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-647-481-2

Query Match 23.1%; Score 479; DB 1; Length 560;
Best Local Similarity 28.2%; Pred. No.2.3e-42;
Matches 112; Conservative 68; Mismatches 155; Indels 62; Gaps 7;

Oy 12 VPSLCISARYGIALVLFHFCNFTTIAQNVINITWAVANVNSTSPQSQINDSEVL PVDSEFG 71
Db 161 IPSARVHYGCVI-----FVRILGLVGEVGYPC----- 191
Oy 72 LSKAKSLPAMKSSILGGCGAIEWRGPPQPERBRLCSIALSGMLGCFYLLIGCFISETL 131
Db 191 -----HGIMSKWAPPLERSRLATTAFCGSYAGVAVVIMPLAGVYVYS 232

```

```

QY 132 GMPFEYTFGGGCGCCCLMFMEVYIDDPYSYPMISSEKEYIISLKQVSSKQ---PL 188
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 223 GMSSTFYVYGGSGITWYLFEMFLVLYSEPALHSSISEEKYI---EDAIGESAKLMP 288
QY 189 -----PIKAMRLSPDPIWSTICGCFSHQWLVTMNVYIPIYISSVYHVNIRDNGLSALPF 243
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 289 TKFSTPMRRFETSMVDVYAIIVANFCRSMTEFYLLISQDPYEFEEVCFEISKVGLV 348
QY 244 IVAMVIGAVGGLADFLPLTKFRRLT-VAKITITLGSLSALIVSLPYLNSYTTATL 302
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 349 LVMITIVIGQIADFLRSRIMSTTNVRLKNNCGGFGMEATLLLVGVSXSGKV-AISF 407
QY 303 LTLSCGSLTLCGSGYIYNVDIAPRYSFLMGASRGSFIAPYIVTVSGFLLSQDPERG 362
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 408 LVLANGFSGFAISGNNVHDLIAPRYASITLKGISNCGVGLSGWCPYIIGAMKHKRTREE 467
QY 363 WRNVFFLLFAVNLGLLFYLLIFGEADVQEMAKERKLT 399
      |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 468 WQYVPLIASLVHYGGVIFYGVPASGEKQWAPRPEKMS 504

RESULT      8
US-08-430-033A-2
; Sequence 2, Application US/08430033A
; Patent No. 5686266
;
GENERAL INFORMATION:
APPLICANT: NI, Bihui
APPLICANT: PAUL, Steven M.
TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/430,033A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: GAYLO, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-10006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 560 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-430-033A-2

Query Match      23.1%; Score 479; DB 1; Length 560;
Best Local Similarity 28.2%; Pred. No. 2.3e-42;
Matches 112; Conservative 68; Mismatches 155; Indels 62; Gaps 7;

QY 12 VPSLCSARYGIALVYHPCFTTIAQNVIMNIMVAMVNSTSPQSQNDSEVLVPVPSFG 71
      |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 161 IPSAARVHYGCVI-----FVRILQGLVGVYTPAC----- 191

72 LSKAPKSP;PAVSSILGGOFAIWERNGPPQERSRSLCSIALSGMLGCFETAILIGGFISETL 131

```

Db 191 -----HGIMSKWAPPLERSRLATFAFCGSYAGAVVAMPLAGLVQYS 232

QY 132 GMPFEYIFGVCVCCCLMPVYIYDDVSYPMWISTSEKEYIISLKQOVSSKQ---PL 188

Db 233 GWSVFYFYGSGIFWYLFMWLVSYESPALHPSISEERKTY---EPAIGESAKLMP 288

QY 189 -----PIKAMRLSLPIWISICLGCFSHOWLVSTMVYIPTYISSYHVNIIRDNGLSALPF 243

Db 289 TKFSFPMRFTSMVYAIYANFCRSMWTFYLLISQPDYEEVGFELSKVGLVSALPH 348

QY 244 IVAWYIGVGTIADFLTKKFRILT-VRKIATITIGSLPSSALIVSLPYLNSGYTTATL 302

Db 349 LVMTIIVIGGIADFLRSRIMSTYNVRKIMNCGFMEATLLIVGYSHSKGV-AISF 407

QY 303 LTLSCGLTLCOSGIYINVLDIAPRYSFIMGASRGFSSIAPIVPTVSGFLSDPPEFG 362

Db 408 LVLAVGFSGFAISGFNVNHLDIAPRYASILMGISNGVETLSGMWCPITIVGAMTKHKTREE 467

QY 363 WRNVFLLFAVNLGLLFYLLIFGEADVOMAKERKLT 399

Db 468 WQYVFLIASLVHYGVIIFYGVAFSGEKQPMAPPEMS 504

RESULT 9

PCT-US96-05792-2

: Sequence 2, Application PC/TUS9605792

: GENERAL INFORMATION:

: APPLICANT: NI, Binhui

: APPLICANT: Paul, Steven M.

: TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC

: TITLE OF INVENTION: PHOSPHATE COTRANSPROTHER

: NUMBER OF SEQUENCES: 3

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Eli Lilly and Company

: STREET: Lilly Corporate Center

: CITY: Indianapolis

: STATE: Indiana

: COUNTRY: United States of America

: ZIP: 46285

: COMPUTER READABLE FORM:

: MEDIUM TYPE: floppy disk

: COMPUTER: IBM PC compatible

: OPERATING SYSTEM: PC-DOS/MS-DOS

: SOFTWARE: Patentin Release #1.0, Version #1.30

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: PCT/US96/05792

: FILING DATE:

: CLASSIFICATION:

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER: 08/430,033

: FILING DATE: April 27, 1995

: ATTORNEY/AGENT INFORMATION:

: NAME: Gaylo, Paul J.

: REGISTRATION NUMBER: 36,808

: REFERENCE/DOCKET NUMBER: X-10006

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: (317) 276-0756

: TELEFAX: (317) 276-3861

: INFORMATION FOR SEQ ID NO: 2:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 560 amino acids

: TYPE: amino acid

: TOPOLOGY: linear

: MOLECULE TYPE: protein

PCT-US96-05792-2

Query Match 23.1%; Score 479; DB 4; Length 560;

Best Local Similarity 28.2%; Pred. No. 2.3e-42;

Matches 112; Conservative 68; Mismatches 155; Indels 62; Gaps 7;

QY 12 VPSICSAHYGIALVHRCNFTIAQNVIMNITVAMVNSTSPQSQINDSSEVLVPDSFGG 71

Db 161 IPSARVHRCYI-----FVRILOGVGYTPAC----- 191

QY 72 LSKADSLPANSILIGGFALWERNRPPQERSRLCSIALSGMLACCTAIIIGFISETL 131

Db 191 -----HGIMSKWAPPLERSRLATFAFCGSYAGAVVAMPLAGLVQYS 232

QY 132 GMPFEYIFGVCVCCCLMPVYIYDDVSYPMWISTSEKEYIISLKQOVSSKQ---PL 188

Db 233 GWSVFYFYGSGIFWYLFMWLVSYESPALHPSISEERKTY---EPAIGESAKLMP 288

QY 189 -----PIKAMRLSLPIWISICLGCFSHOWLVSTMVYIPTYISSYHVNIIRDNGLSALPF 243

Db 289 TKFSFPMRFTSMVYAIYANFCRSMWTFYLLISQPDYEEVGFELSKVGLVSALPH 348

QY 244 IVAWYIGVGTIADFLTKKFRILT-VRKIATITIGSLPSSALIVSLPYLNSGYTTATL 302

Db 349 LVMTIIVIGGIADFLRSRIMSTYNVRKIMNCGFMEATLLIVGYSHSKGV-AISF 407

QY 303 LTLSCGLTLCOSGIYINVLDIAPRYSFIMGASRGFSSIAPIVPTVSGFLSDPPEFG 362

Db 408 LVLAVGFSGFAISGFNVNHLDIAPRYASILMGISNGVETLSGMWCPITIVGAMTKHKTREE 467

QY 363 WRNVFLLFAVNLGLLFYLLIFGEADVOMAKERKLT 399

Db 468 WQYVFLIASLVHYGVIIFYGVAFSGEKQPMAPPEMS 504

RESULT 10

US-08-805-118-4

: Sequence 4, Application US/08805118

: Patent No. 5985604

: GENERAL INFORMATION:

: APPLICANT: Lal, Preeti

: APPLICANT: Bandman, Olga

: TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT

: TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER

: NUMBER OF SEQUENCES: 7

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Incyte Pharmaceuticals, Inc.

: STREET: 3174 Porter Drive

: CITY: Palo Alto

: STATE: CA

: COUNTRY: US

: ZIP: 94304

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Diskette

: COMPUTER: IBM compatible

: OPERATING SYSTEM: DOS

: SOFTWARE: FastSeq Version 2.0

: CURRENT APPLICATION DATA:

: APPLICATION NUMBER: US/08/805,118

: FILING DATE: Filed Herewith

: CLASSIFICATION: 435

: PRIOR APPLICATION DATA:

: APPLICATION NUMBER:

: FILING DATE:

: ATTORNEY/AGENT INFORMATION:

: NAME: Billings, Lucy J.

: REGISTRATION NUMBER: 36,749

: REFERENCE/DOCKET NUMBER: PF-0221 US

: TELECOMMUNICATION INFORMATION:

: TELEPHONE: 415-855-0555

: TELEFAX: 415-845-4166

: INFORMATION FOR SEQ ID NO: 4:

: SEQUENCE CHARACTERISTICS:

: LENGTH: 560 amino acids

: TYPE: amino acid

: STRANDEDNESS: single

: TOPOLOGY: linear

: IMMEDIATE SOURCE:

: LIBRARY: GenBank

CLONE: 507415

US-08-805-118-4

Query Match 22.9%; Score 475; DB 2; Length 560;
Best Local Similarity 28.0%; Pred. No. 6,1e-42;
Matches 111; Conservative 66; Mismatches 158; Indels 62; Gaps 6;

QY 12 VPSLCSARGIALVLFHFCFTTIAONVIMINIMVAVNSTSPQSLDNSEVLVPVDSFG 71
DB 161 IPSARVHKGCVI-----FVRILOGIVEGVTPAC----- 191
QY 72 LSKARKSLPAKSSILGGOFAPWERNRGPPOERSRLCSIALSGMLGCTAILIGGFSETL 131
DB 191 -----HGWSKNAPLERSLATTAFCGAYAGVAVMLAGLVQYS 232
QY 132 GMPFPIYFGVGVCCCLMFVYIYDDPVSPWISTSEKEYIISLKQOVSS----- 185
DB 233 GMSVFEYVYGSFGIFWYLFMLVSESPALHPSISEERKVI---EDAHGSAKLMPV 288
QY 185 -KQPLPIKMLRSLPWSICLCGFSHOWLVTMVVYIPYIISVYHVNRDGLSLALPF 243
DB 289 TKFNPWRRFPFSPMPYALITVANFCRSMTFYLLISQPAFEEVEGFEESKVGLSALPH 348
QY 244 IVAMVIGWGYLADFLTKKRLIT-VKRIATILGSLPSALIVSLPYLNGYITATL 302
DB 349 LVMTIIVPIGGIADFLRSRHMSTTNVRKLMNCGFGMEATLLLVG:SHSKGV-AISF 407
QY 303 LTLSCGLSTLCOSGIYINVDIAPRYSFLMGASRGFSSIAPIVPTVSGFLLSODPERG 362
DB 408 LVLAAGFSGFAISGFVNHLDIAPRYASILMGISNGVGLSCVCPILVIGAMTKHKTRE 467
QY 363 WRNVFELFAVNLGLFLFLIGLEADVOEMAKERKIT 399
DB 468 WQVFLASLVHGYGVIFGVFASGEKQPMAPPEEMS 504

RESULT 11

US-08-898-976-2

; Sequence 2, Application US/08898976
; Patent No. 5891670
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin
; APPLICANT: Lonetto, Michael
; APPLICANT: Warren, Patrick
; TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE
; NUMBER OF INVENTION: 6
; CORRESPONDENCE ADDRESS:
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/898,976
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10044
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259

; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 462 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

US-08-898-976-2

Query Match 5.9%; Score 122.5; DB 2; Length 462;
Best Local Similarity 17.7%; Pred. No. 0.0001;
Matches 76; Conservative 86; Mismatches 158; Indels 109; Gaps 18;

QY 20 YGIALVLFHFCFTTIAONVIMN---ITVAVANSTSPQSLDNSEVLVPVDSFGSLKAP 76
DB 92 YGVILIT---FGSIIGNIFQHSPLTLVGRITQTA-----GLAAAE 129
QY 77 KSLPAKSSILGGOFAPWERNRGPPOERSRLCSIALSGMLGCTAILIGGFSETLGM-- 134
DB 130 T-----LVYIVAKYLSKEDQKTYLGISTSSYLSLYIGTSLGSGFIPTTYLHMTN 178
QY 134 -----PFVEYI-----FGVGCVCCLMFVYIYDDPVSPWISTS 168
DB 179 MELIALIVFTLPFLKLPKKNNTKKAHLDFYGLLVATITATWLF--ITNFMNL---- 234
QY 169 EKEYIISL-----KQOVSSKQPLPIKMLRSLPWSICLCGFSHOWLVTMVVYIPT 222
DB 234 ---YMGALIAIIVFALYIKNAORPLVNRKSFQNKRYASFLFVFWVAIQLOIYTPFP 290
QY 223 YISSVHVHNRDGLSALPFIYAMVIGWGYLADFLTKKRLITVKRIATILGSLPS 282
DB 291 IMEQIYHLDPTTSLLVPGYIVAVIVGALSGIGEYLSKQAIITATILALST-ILPA 349
QY 283 -----SALIVSLPYLNGYITATLTLSCGLSTLCOSGIYINVDIAPRYSFLMGA 335
DB 350 FAVGNHISTFVISMIFPAGSFALMTAPL-LNEAIKT-----IDLNFTGVALIGYNIITVY 403
QY 336 SNGFSSIAPIVPTVSGFLLSODP-----EGWNRVFLFAVNLGLFLFLIGLEAD 388
DB 404 A---VSVGIALIAAALIDFALNPGDALSHPG--IILIIILGMSIVGLVLEVI----- 454
QY 389 VOEMAKER 397
DB 454 LNRWTOSEK 462

RESULT 12

US-08-898-976-4

; Sequence 4, Application US/08898976
; Patent No. 5891670
; GENERAL INFORMATION:
; APPLICANT: Burnham, Martin
; APPLICANT: Lonetto, Michael
; APPLICANT: Warren, Patrick
; TITLE OF INVENTION: NOVEL TETRACYCLINE RESISTANCE
; NUMBER OF INVENTION: 6
; CORRESPONDENCE ADDRESS:
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/898,976
; FILING DATE:

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Bloom, Allen
REGISTRATION NUMBER: 29,135
REFERENCE/DOCKET NUMBER: GM10044
TELECOMMUNICATION INFORMATION:
TELEPHONE: 609-520-3214
TELEFAX: 609-520-3259
TELEX:
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 462 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-898-976-4

Query Match 5.9%; Score 122.5; DB 2; Length 462;
Best Local Similarity 17.7%; Pred. No. 0.0001;
Matches 76; Conservative 86; Mismatches 158; Indels 109; Gaps 18;

QY 20 YGAIYALHFCNFTTIAONVIM--ITWYAVNSTSPQSOLNDSSEVLVDSFGGSKAP 76
DB 92 YGVILII-----FGSTIGNFQHSPLTLVGRIGTA-----GLAANE 129
QY 77 KSLPKKSSILGQFALW-ERMGPPOBRSLCSIALSGMLGCTALLIGFTSETLW-- 134
DB 130 T-----LVYIYAKYLSKEDOKTYLGLSTSYSLSLVIGTSLGCFYFLHMTN 178
QY 134 -----PEVEYI-----FGVGVCCOLMFVYIYDDPSYVWIS 168
DB 179 MFLILIVFTLPFLFKLLPKENNTNKAHLDFVGLLVATTATVWLP--ITNFWNL--- 234
QY 169 EKEYIISL-----KQVGSCKOPLPKAMLRSLPWSICIGCFSHQWLVSTWVYIPT 222
DB 234 ---YMGALIAIIVFALYIKNAORPLVKNKSPQNKRYASFLFIVYMAIQGYFTTFP 290
QY 223 YISSYHNINRNGLSLAPFIVAVNIGWGYLADFLITKFKRLLITVTKATITGLSPS 282
DB 291 IMEOLYHQLDTSLILPGYIVAVYVGLSGKIGEYLNKQAITAILIALSL-ILPA 349
QY 283 -----SALIVSLPYLNSGYITATALLTSCGLSTLQSGIYINVDIARRYSFLMGA 335
DB 350 FAVGNHISIFVISMTEFFAGSFAIMAPL-LNEAIKT-----IDLMTGVAIGFVNLITN 403
QY 336 SRGSSIAPIVIVPYVSGFLSQDP-----EFGWRNVFLLFAVNLGLLFYILFGEAD 388
DB 404 A--VSVGIAIAALIDFKALNFPNDALSHFG--ILITLIGLSITGVLFVY----- 454
QY 389 VOENAKERK 397
DB 454 LNRRTQSEK 462

RESULT 13
US-08-466-033-15
Sequence 15, Application US/08466033
Patent No. 5766840
GENERAL INFORMATION:
APPLICANT: Kim, Jungsuh P.
APPLICANT: Wages, John
APPLICANT: Young, Lavonne M.
APPLICANT: Ely, Kirk E.
APPLICANT: Linnen, Jeffrey M.
TITLE OF INVENTION: Hepatitis G Virus and Molecular
NUMBER OF SEQUENCES: 277
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates

STREET: 350 Cambridge Ave., Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,033
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/389,886
FILING DATE: 15-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/357,509
FILING DATE: 16-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,729
FILING DATE: 26-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/344,271
FILING DATE: 23-NOV-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,558
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/285,543
FILING DATE: 03-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/246,985
FILING DATE: 20-MAY-1994
ATTORNEY/AGENT INFORMATION:
NAME: Fabian, Gary R.
REGISTRATION NUMBER: 33,875
REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 324-0880
TELEFAX: (415) 324-0960
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 2873 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-466-033-15

Query Match 5.1%; Score 106.5; DB 1; Length 2873;
Best Local Similarity 22.9%; Pred. No. 0.078;
Matches 89; Conservative 42; Mismatches 133; Indels 125; Gaps 22;

QY 79 LPAKSTILG-QGFALMERMGPOBRSLCSIALSGMLGCTALLIGFTSET-----L 131
DB 637 LPVYSMTLGLANLVYLRWLGPO--RLMFLVLMKLRAGAPPLALMG-ISARGRGTSVL 692
QY 132 GWPEVF-----YIRGGVCCOLMFVYIYDDPSYVW-----ISSEKEYIISL 177
DB 693 GAERCFDATEREVDPSVYLGWV-VASVYMAIALLSMSAGGWRKRAVIYRTWCKGY--QAT 749
QY 178 KQVYGSCKOPL-----PKAMLRSLPWSICIGCFSHQWLVSTWVYIPTVYISSYHNIR 233
DB 750 RORV--VRSPLEGGRPAKPL-----TFAMCLA--SYIMPDVAVMVVVALVLF----- 794
QY 234 DNGLSALPFTVAMVIGWGYLADFLITKFK--RLITVAKKITITIGSLPSSALIVSLPY 291
DB 794 --GLFDALD---WA-----LEEITVSRPSLRRLARVECCVMAGERATVRLVSKMC 840
QY 292 LNSGY-----ITATALLTSCGLSTLQSGI 317

DB 841 ARGAVLEFDHMGSPRAVAKERLLEMDALEPLSFRTDRCRIIRDAARTLSCG---QCYMGL 897
QY 318 YINVLDIAPRYSSEFLMGASRGFSSIAPIVTPV-----SGF-----LISQDEFG 362
DB 898 PV-----VARRGDEVLLIGFODVNHLPFGFPTAPVPIRRCGKGLGVTKAALGRDPDLH 953
QY 363 WRNVFFLLFAVN-----LLGLLFYLLFG 385
DB 954 PGNVVVLGTATSRSMGTCLNGLLFTTFHG 982

RESULT 14
US-08-638-911A-2
; Sequence 2, Application US/08638911A
; Patent No. 5766916
; GENERAL INFORMATION:
; APPLICANT: Belayev, Alexander S.
; APPLICANT: Chong, Susan M.
; TITLE OF INVENTION: Hepatitis G Virus Protease
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/638,911A
; FILING DATE: 04/24/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Evans, Susan T.
; REGISTRATION NUMBER: 38,443
; REFERENCE/DOCKET NUMBER: 4600-0124
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2873 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-638-911A-2

Query Match 5.1%; Score 106.5; DB 1; Length 2873;
Best Local Similarity 22.9%; Pred. No. 0.078;
Matches 89; Conservative 42; Mismatches 133; Indels 125; Gaps 22;

QY 79 LPAKSSILG-GQFAIWERKCPPOERSRLCSIALSGMLGCFYAILIGGFISER-----L 131
DB 637 LPVYASHILGLANLVYFRMLGPQ--RLMFLVLMKLAGAFPLALMG-ISAIRGRYSVL 692
QY 132 GMPFVF-----YIFGGGVCVCLLMFVYIYDDPSYPM-----ISTSEKREYISSL 177
DB 693 GAFCFCDATEVETSVLGVV-VASVYAMAIALSSMSAGWRKKAIVYRTWCKGY--QAI 749
QY 178 KQGVSSSKQPL-----PIKAMLRSLPIWISICLGCFSHOWLVTWVYIPIYISSVYHVNIR 233
DB 750 ROHYV--VRSPLGGRPAKPL-----TFAWCLA--SYIMPDVAVMVVVVALLVLF----- 794
QY 234 DNGLLSALPPIVAVMVGWGYLADFLITKFF--RLITYAKITILGSLPSSALIYSLPY 291
DB 794 --GLEFDALD---WA-----LEEILVSRPSLRRLARAVECCVMGEKATYTVRLSKMC 840
QY 292 LNSGY-----ITTATALLTLSCGLSTIQQSGI 317

DB 841 ARGAVLEFDHMGSPRAVAKERLLEMDALEPLSFRTDRCRIIRDAARTLSCG---QCYMGL 897
QY 318 YINVLDIAPRYSSEFLMGASRGFSSIAPIVTPV-----SGF-----LISQDEFG 362
DB 898 PV-----VARRGDEVLLIGFODVNHLPFGFPTAPVPIRRCGKGLGVTKAALGRDPDLH 953
QY 363 WRNVFFLLFAVN-----LLGLLFYLLFG 385
DB 954 PGNVVVLGTATSRSMGTCLNGLLFTTFHG 982

RESULT 15
US-08-444-733-15
; Sequence 15, Application US/08444733
; Patent No. 5824507
; GENERAL INFORMATION:
; APPLICANT: Kim, Jungsuh P.
; APPLICANT: Wages, John
; APPLICANT: Young, Lavonne M.
; APPLICANT: Fry, Kirk E.
; APPLICANT: Linen, Jeffrey M.
; TITLE OF INVENTION: Hepatitis G Virus and Molecular
; NUMBER OF SEQUENCES: 277
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave., Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,733
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/389,886
; FILING DATE: 15-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/357,509
; FILING DATE: 16-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,729
; FILING DATE: 26-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,271
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/285,558
; FILING DATE: 03-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/246,985
; FILING DATE: 20-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Fabian, Gary R.
; REGISTRATION NUMBER: 33,875
; REFERENCE/DOCKET NUMBER: 4600-0201.36/G100P11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2873 amino acids
; TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-444-733-15

Query Match 5.1%; Score 106.5; DB 2; Length 2873;
Best Local Similarity 22.9%; Pred. No. 0.078;
Matches 89; Conservative 42; Mismatches 133; Indels 125; Gaps 22;

```

QY 79 LPAKSSILG-GQFAIWRMGPPQERSRLCSIALSGMLIGCFATILIGFISST-----L 131
   || || || || || || || || || || || || || || || || || || || || ||
Db 637 LPVYSMLIGIANLYLFRNLGPP---RLMFVLMKLARGAFPLALMG-ISAIRGRTSVL 692
   || || || || || || || || || || || || || || || || || || || || ||
QY 132 GMPFVF-----YIFGVGCVCCLMFVVIYDDPVSYPW-----ISTSEKVIISSL 177
   || || || || || || || || || || || || || || || || || || || || ||
Db 693 GAFFCFDATERVDTSVLGMV-VASVYAMAIALISSMSAGMRHKAVIYRTWCKGY--QAI 749
   || || || || || || || || || || || || || || || || || || || || ||
QY 178 KOVGSKKQPL---PIKAMLRSLPIWSTICGCFSHQMLVSTMVYIIFTYISSVYHVNIR 233
   || || || || || || || || || || || || || || || || || || || || ||
Db 750 RQRY--VRSPLEGGRPAKPL-----TFAMCLA--SYIMPDAMVMVVALVLF----- 794
   || || || || || || || || || || || || || || || || || || || || ||
QY 234 DNGLLSALPFIYAWVIGWGYLADFLTKKF--RLITVRKIATILIGSLPSSALIVSLPY 291
   || || || || || || || || || || || || || || || || || || || || ||
Db 794 --GLFDALD---WA-----LEELIVSRPSLRRLARVYECVMAGEKATTVRLVSKMC 840
   || || || || || || || || || || || || || || || || || || || || ||
QY 292 LNSGY-----ITATALLTLSCGLSTLCQSGI 317
   || || || || || || || || || || || || || || || || || || || || ||
Db 841 ARGAYLFDHMGFSRAVKEKRLLEMDALEPLSFTRDCRIIRDAARTLSCG---QCYMGL 897
   || || || || || || || || || || || || || || || || || || || || ||
QY 318 YINVLDIAPRYSSFLMGASRGFSIAPYIVPV-----SGF-----LLSODPERG 362
   || || || || || || || || || || || || || || || || || || || || ||
Db 898 PY-----VARRGDEVILGVFODVNHLPPGFVPTAPAVYIRRCGKGLGVTAKAALTGRDPDLH 953
   || || || || || || || || || || || || || || || || || || || || ||
QY 363 WRNVEFLFAVN-----LLGLLFYLIIG 385
   || || || || || || || || || || || || || || || || || || || || ||
Db 954 PGNVAVLGTATSRSMGTCINGLLFTTFPHG 982
   || || || || || || || || || || || || || || || || || || || || ||

```

Search completed: May 25, 2000, 14:25:01
Job time: 2804 sec

OM of: US-09-391-958-1 to: GenEmbl.* out_format : pfs
Date: May 25, 2000 1:47 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODEL-frame+pin.model -DEV-x1p
-O/cgcn2_1/USPRO.spool/US09391958/runat.24052000_213330_7161/app-query.fasta.1
-DB-GenEmbl -OPMT-fastlap -SUFFIX=rgc -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPEL=0.000 -LOOPEXT=0.500 -GAPOP=4.500
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
-DELEXT=7.000 -START=1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=PCT -ALIGN=15 -MODE=LOCAL
-OUTFM=pfs -NORM-ext -MINLEN=0 -MAXLEN=1000000 -USBR=US09391958
-NCPU=6 -ICPU=3 -NO_XLIPX -WAIT -THREADS=1

Search information block:

Query: US-09-391-958-1
Query Length: 401
Database: GenEmbl.*
Database sequences: 882769
Database length: -486395729
Search time (sec): 737.170000

score_list:

Sequence	Strd Orig	ZScore	EScore	len	Documentation
gb_pat:AR036571	+ 2061.00	3007.97	1.7e-159	1780	AR036571 Sequence 19 from pate
gb_p3:HSU90545	+ 2061.00	3007.90	1.7e-159	1795	U90545 Human sodium phosphate
gb_p1:AB020527	+ 1023.50	1483.30	1.4e-74	2626	AB020527 Homo sapiens mRNA for sodi
gb_p2:HSNAB1	+ 964.00	1400.99	5.6e-70	1549	X71355 H.sapiens mRNA for sodi
gb_ro:RNU28504	+ 959.00	1392.80	1.6e-69	1700	U78504 Rattus norvegicus Na+/H
gb_ro:NMNPT1CT	+ 951.00	1380.10	8.1e-69	1885	X77241 M.musculus Npt1 mRNA fo
gb_p1:HMARP14	+ 944.00	1370.30	2.8e-68	1794	D28522 Human mRNA for renal Na
gb_com:RAMRCCTX	+ 937.00	1359.73	1.1e-67	1855	X76466 Rabbit renal cortical N
gb_pat:AR036570	+ 932.00	1350.53	3.6e-67	2266	AR036570 Sequence 18 from pate
gb_p3:HSU90544	+ 932.00	1350.47	3.6e-67	2281	U90544 Human sodium phosphate
gb_hcg2:AL1138726	+ 876.50	1230.66	1.7e-60	140266	AL1138726 Homo sapiens chromo
gb_p3:HS283747	+ 615.50	889.56	2.9e-41	2512	AJ387747 Homo sapiens mRNA for
gb_p2:HS283953	+ 595.00	869.35	2.3e-40	570	I283953 H.sapiens mRNA; clone CD
gb_pat:140028	+ 479.00	684.71	4.4e-30	2716	I40028 Sequence 1 from patent
gb_pat:140029	+ 479.00	684.71	4.4e-30	2716	I40029 Sequence 3 from patent
gb_pat:140220	+ 479.00	684.71	4.4e-30	2716	I40220 Sequence 1 from patent
gb_pat:140221	+ 479.00	684.71	4.4e-30	2716	I40221 Sequence 3 from patent
gb_pat:173259	+ 479.00	684.71	4.4e-30	2716	I73259 Sequence 1 from patent
gb_pat:173260	+ 479.00	684.71	4.4e-30	2716	I73260 Sequence 3 from patent
gb_ro:RNU07609	+ 475.00	681.59	6.5e-30	2024	U07609 Rattus norvegicus brain
gb_in2:AF095787	+ 466.50	668.44	3.5e-29	2178	AF095787 Caenorhabditis elegans
gb_hcg4:AC014246	+ 445.50	605.12	4.8e-26	33444	AC014246 Drosophila melanog
gb_hcg6:AC020077	+ 443.50	612.97	1.1e-25	47723	AC020077 Drosophila melanog
gb_in2:AF024691	+ 440.50	616.38	2.8e-26	9703	AF024691 Drosophila ananassa
gb_hcg4:AC014600	+ 430.50	589.96	8.3e-25	34229	AC014600 Drosophila melanog
em_in:AC003924	+ 430.50	581.73	2.4e-24	82692	AC003924 Drosophila melanog
gb_in2:AC006302	+ 430.50	576.86	4.5e-24	139238	AC006302 Drosophila melanog
gb_hcg4:AC013869	+ 419.50	583.96	1.8e-24	11568	AC013869 Drosophila melanog
gb_hcg4:AC015336	+ 417.00	585.24	1.5e-24	6806	AC015336 Drosophila melanog
gb_hcg12:AC009391	+ 415.00	549.51	1.5e-22	228783	AC009391 Drosophila melanog
gb_in2:AC004345	+ 410.00	551.26	1.2e-22	86419	AC004345 Drosophila melanog
gb_hcg5:AC018326	+ 410.00	548.97	1.6e-22	110482	AC018326 Drosophila melanog
gb_in2:AC004246	+ 407.50	552.92	1.0e-22	50279	AC004246 Drosophila melanog
gb_hcg6:AC019866	+ 407.50	548.84	1.6e-22	75607	AC019866 Drosophila melanog
gb_in1:DMNACOTR	+ 404.50	575.63	5.2e-24	2674	I07720 D.melanogaster mRNA for
gb_hcg6:AC019757	+ 400.50	539.81	5.2e-22	66241	AC019757 Drosophila melanog
gb_hcg6:AC020165	+ 400.00	562.26	2.9e-23	5527	AC020165 Drosophila melanog
gb_hcg6:AC020123	+ 400.00	539.92	3.1e-22	60528	AC020123 Drosophila melanog
gb_hcg5:AC008232	+ 397.00	529.02	2.1e-21	121584	AC008232 Drosophila melanog
gb_hcg5:AC005463	+ 390.00	522.02	5.1e-21	85651	AC005463 Drosophila melanog
gb_in2:AC005448	+ 390.00	521.89	5.1e-21	86889	AC005448 Drosophila melanog
gb_hcg11:AC007084	+ 390.00	517.52	9.0e-21	138793	AC007084 Drosophila melanog

gb_hcg8:AC011065 + 387.50 513.94 1.4e-20 137525 I AC011065 Drosophila melano
gb_hcg6:AC020394 + 387.00 525.21 3.4e-21 37983 I AC020394 Drosophila melano
gb_in1:CEC38C10 - 381.50 518.13 8.3e-21 34193 I Z19153 Caenorhabditis eleg

seq_name: gb_pat:AR036571

seq documentation block:

LOCUS	AR036571	1780 bp	DNA	PAT	29-SEP-1999
DEFINITION	Sequence 19 from patent US 5872237.				
ACCESSION	AR036571				
VERSION	AR036571.1	GI:5953239			
KEYWORDS					
SOURCE	Unknown.				
ORGANISM	Unknown.				

REFERENCE
1 (bases 1 to 1780)

Feder, J.Nachman, Kromal, G.Scott, Lauer, P.M., Ruddy, D.A., Thomas, W.,
Tsushinashi, Z. and Wolff, R.K.

MEGABASE TRANSCRIPT MAP: novel sequences and antibodies thereto
Patent: US 5872237-A 19 16-FEB-1999;

TITLE
JOURNAL
FEATURES
source
Location/Qualifiers
1..1780

BASE COUNT
ORIGIN
458 a 436 c 402 g 484 t

alignment_scores:

Quality: 2061.00 Length: 401
Ratio: 5.165 Gaps: 0
Percent Similarity: 99.501 Percent Identity: 99.252

alignment_block:
US-09-391-958-1 x AR036571 ..

Align seg 1/1 to: AR036571 from: 1 to: 1780

1	MetlnValAspGluThrLeuIleProArgLysValProSerLeuGlyCysE	17
377	ATGCAGATGTGATGAGACATGATCCCGAGAAAGCTCCAGTTATGTTC	426
17	TAATATGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG	34
427	TGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC	476
34	LeaIaGlnAsnValIleLeuMetValIleMetValIleMetValIleSer	50
477	TAGCACAATATGTCATCATGAAATCAATGATGATGATGATGATGATG	526
51	TherProGlnSerGlnLeuAsnAspSerSerGlnValIleProValAs	67
527	ACAAGCCCTCAATCCAGCTCAATGATTCCTGAGTGTGCTGCTGGA	576
67	PserPheGlyLeuSerSerLysAlaProLysSerLeuProAlaLysSer	84
577	CTCATTTGTCGTCCTAAGTAAAGCCCAAGAGCTTCTCTCAAGTCTT	626
84	erileLeuGlyGlnPheAlaIleTrpGluArgTrpGlyProProGln	100
627	CAATATCTGGGGGTCAGTTTCAATTTGGAAAGTGGGGCCCTCCACA	676
101	GluArgSerArgLeuGlySerIleAlaLeuSerGlyMetLeuLeuGly	117
117	sPheThrAlaIleLeuIleGlyGlyPheIleSerGlnThrLeuGlyTrp	134
727	CTTTACTGCTCATCTCATAGTGGCTTCATTAGTGAACCTTGGGTGC	776
134	ropheValPheTrpIlePheGlyGlyValGlyCysValCysLeuLeu	150
777	CCCTTGTCTCTATATATCTTGTGAGAGTGTGCTGTGCTGCTTCTC	826

```

151 TrpPheValValIleTyraSPaSPProValSerTyrrProTrIleSerTh 167
|||||
827 TGGTTTGTGGATTATGATGATGACCCCTTTCTATCCATGATGATAGAC 876
|||||
167 rSerGluGluGluTyrIleIleSerSerIleuGluGluValIleSers 184
|||||
877 CTCGAAAAAGAAATACATATCTCTCTGAAACAAACAGTCGGTCTT 926
|||||
184 eIryGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTirp 200
|||||
927 CTAGACAGCCCTTCCCATCAACATATGCTCATCTCATCCATTTGG 976
|||||
201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThMetVa 217
|||||
977 TCATATATGTTAGGCTGTTTCAGCCATCAATGTTAGTACACAAATGT 1026
|||||
217 ValTyrrIleProThTyrrIleSerSerValTyrrHisValAsnIleArg 234
|||||
1027 TGTATACATACCAACTTACATACAGCTCTGTGTACCATGTTAACATCAG 1076
|||||
234 sPaSngIleLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
|||||
1077 ACAATGACACTTATCTGACCTTCTTATGTTGCTGGGTCAATAGGC 1126
|||||
251 MetValGlyGlyTyrrLeuAlaAspPheLeuThrlYsIysPheArgIe 267
|||||
1127 ATGGTGGAGGCTATCTGGCAGATTCCTTCTTACCAAAAAAGTTTACACT 1176
|||||
267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSerA 284
|||||
1177 CATCAGCTGTAGGAAATTCGCCAATTTTAGGAAGTCTCCCTCTTCAG 1226
|||||
284 lAleuIleValSerLeuProTyrrLeuAsnSerGlyTyrrIleThrAlaTh 300
|||||
1227 CACTCATTTGCTCTGCTTACCTCAATTCGGCTATATACAGCAACT 1276
|||||
301 AlaLeuLeuThrlLeuSerCysGlyLeuSerThrlLeuCysGlnSerGly 317
|||||
1277 GCCTTCTGACGCTCTCTGCGGATTACGACATTTGTGTACAGCAGAGAT 1326
|||||
317 eTyrrIleAsnValIleuAspIleAlaProArgTyrrSerSerPheLeuMet 334
|||||
1327 TTATATACATATGCTTATGATATTGCTCCAGATTTTCCAGTTTCTCATGG 1376
|||||
334 lYAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
|||||
1377 GAGCATCAAGAGATTTTCAGCATAGCACCTGTCAATTGACCACTGTC 1426
|||||
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValPh 367
|||||
1427 ACGGATTTCTTCTTAGTCAGACCTGAGTTTGGGTGAGGAAATGTCTT 1476
|||||
367 ePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrrIleIleP 384
|||||
1477 CTTCTTGTGCTTGTGCGCTTAACCTGTAGACTACTCTTCACTCATATAT 1526
|||||
384 heGlyGluAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrArg 400
|||||
1527 TTGGAGAACAGATGTCACAAAGATGGCTTAAGAGAGAAACTCACTCGT 1576
|||||
401 leu 401
|||
1577 TTA 1579

```

seq_name: gb_pr3:HSU90545

seq_documentation_block:

LOCUS HSU90545 1795 bp mRNA PRI 02-MAY-1997
DEFINITION Human sodium phosphate transporter (NPT4) mRNA, complete cds.
ACCESSION U90545
VERSION U90545.1 GI:2062691
KEYWORDS
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS Ruddy,D.A., Krommal,G.S., Lee,V.K., Mintler,G.A., Quintana,L.,
1 (bases 1 to 1795)
Domingo,R. Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Maps,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchinashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
A 1.1 megabase transcript map of the human hereditary
hemochromatosis locus
JOURNAL
REFERENCE 2 (bases 1 to 1795)
AUTHORS Ruddy,D.A., Krommal,G.S., Lee,V.K., Mintler,G.A., Quintana,L.,
Domingo,R. Jr., Meyer,N.C., Basava,A., McClelland,E., Fullan,A.,
Maps,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchinashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
Direct Submission
Submitted (25-FEB-1997) Sequencing, Mercator Genetics, 4040
Campbell Avenue, Menlo Park, CA 94025, USA
FEATURES
source
1..1795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6p21.3"
377..1582
/gene="NPT4"
377..1582
/gene="NPT4"
/codon_start=1
/product="sodium phosphate transporter"
/protein_id="AA053423.1"
/db_xref="GI:2062692"
/translation="MOVDELILPRKGSLSARGLALVLFNCFTTIAQNVINITM
VAVNVSPOSQNDSEVLVDSFGSLSRAPKSSILGGQFAIMKGPQDR
SLCSIALSGMLGCFPAIILGFISETLGPFFVYFEGGVCYCCILMVIYDDDF
SYPISTSEKEYIISLKOOGVSSKQPLPIKAMRSIPISICGCSRHQMVISTWY
YIPYISSVYHVNIRDNGILSAIPFIYAWYIGWGYGLADFLTKRFLTKRIATY
IGSPSSALIVSLPYLNSGYTTATLALTLSCGSTLQSGIYINVDIARYSEFLMG
ASRGSSTIAPYIVTVSGFLLSDPFEGWNRVFLFAVNLGLLFYILFGEDVQVM
AKERRIRL"

BASE COUNT 473 a 436 c 402 g 484 t
ORIGIN
alignment_scores:
Quality: 2061.00 Length: 401
Ratio: 5.165 Gaps: 0
Percent Similarity: 99.501 Percent Identity: 99.252

alignment_block:

US-09-391-958-1 x HSU90545 ..

Align seg 1/1 to: HSU90545 from: 1 to: 1795

```

1 MetGlnValAspGluThrLeuIleProArgLysValProSerLeuCysE 17
|||||
377 ATGCAAGTGGATGACACACTGATCCCGAGAAAGTCCAAAGTTATGTTTC 426
|||||
17 TAlaArgTyrrGlyIleAlaLeuValLeuHisPheCysAsnPheThrThi 34
|||||
427 TGCTCGCTATGGAATAGCCCTCGTCTTACATTTCTGCAATTTCAACAAGA 476
|||||
34 lAlaGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
|||||
477 TAGCACAATAATGTCAATCATGAAACATCACCATGCTAGCCATGCTCAACAGC 526
|||||
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
|||||
527 ACAAGCCCTCAATCCACAGCTCAATATATCTCTGAGGTCGCTGTGA 576
|||||
67 pSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSers 84
|||||

```

```
577 CTCATTGTGGCCTAAGTAAAGCCCAAGAGCTTCTGCAAAAGTCC 626
84 ertleuaglyglnphealaletrpgluatrglyproproglin 100
|||||
627 CAACTGTTGGGGTCAGTTTGGCAATTTGGGAAAGTGGGGCCCTCCACAA 676
101 gluatrglyserleucysserillealeuSerglymetleuenglycy 117
|||||
677 GAAGCAAGCAGACTCTGACACTTCTTTATCAGGAATGTTACTGGGATG 726
117 spherhAlaleuileuglypheilleSerGluThrleuglytrp 134
|||||
727 CTTTACTGCATCCTCATAGGTGGCTTCATTAGTGAACACCTTGGGAGC 776
134 ropheValpheTrillepheglylyValglycysValglycysleu 150
|||||
777 CTTTGTCTTATATCTTGGAGGTGTGGCTGTGCTGCTGCTCTC 826
151 TrpPheValValleTyraSPAPProValSerTyProTrpIleSerTa 167
|||||
827 TGGTTTGTGTGATTATGATGACCCCTTTTCTTATCATTGATGATNAC 876
167 rSerGluLysgluTryleileSerSerleuLysglnValglysers 184
|||||
877 CTCGAAAAAGAAATACATCATCTCTGAAACACAGGTGGGCTCT 926
184 ertysglnProleuproilleLysAlamelleuarSerleuproiletrp 200
|||||
927 CTAGAGCGCTCTTCCATCAAACTATGCTCAGATCTTACCATTTG 976
201 SerIleCysleuLysPheSerHisGlnTrpValSerThrmetVa 217
|||||
977 TCCATATGTTTAGCTCTTTCAGCATCATAGTTAGTACACAAAGT 1026
217 lValTrilleProThyTrilleSerSerValTyHisValasnllearg 234
|||||
1027 TGTATACATACCACTTACATCAGCTGTGTACATGTTAACTACAG 1076
234 sPasnGlyleuLeuSerAlaleuPropheilleValAlatrPValillegly 250
|||||
1077 ACAATGACCTTATCTGCCCCCTTCTTTATGTTGGCTGGGCTATAG 1126
251 MetValIlglyTyrlleuAlaaspPheleuThrlyslsPheargle 267
|||||
1127 ATGTGGAGAGCTATCTGGCAGATTTCTTCTTCAACCAAAAGTTAG 1176
267 uIleThrValArgLysIleAlaThrIleleuglySerleuProSerSera 284
|||||
1177 CAACTGCTGAGGAAATTTGCCAATTTTAGAGAGTCTCCCTCTCAG 1226
284 lAleuileValSerleuProTyrlleuansSerGlyTyrlleThrAlaThr 300
|||||
1227 CACTCATGTTGCTCTGCTTACTCAATTCGGCTATATCACAGCAACT 1276
301 AlAleuThrleuSerCysGlyleuSerThrleuCysGlnSerGlyI 317
|||||
1277 GCCTGTGTGAGGCTCTCTGGGGATTAGCACATTTGTGTCAGTAGGAT 1326
317 eTyrlleasnValleuAspIleAlaProArgTySerSerPheleuMetG 334
|||||
1327 TTTATATCAATGTTTATATATGCTCCAGGATATTCAGTTTCTCATGG 1376
334 lYAlaSerArgglyPheSerSerIleAlaProValIleValProThrVal 350
|||||
1377 GAGATCAAGAGGATTTTCGAGCATACCTGTCTTATACCACTGTC 1426
351 SerGlyPheleuSerGlnasPProGluPheglyTrpArganValph 367
|||||
1427 AGCGGATTTCTTCTTACTCAGAGCCCTGAGTTTGGTGAGGAAATGCT 1476
367 ePheleuLeuPheAlaValasnlleuLysleuLysPheTyrlleuLeu 384
|||||
1477 CTCTCTGCTGTTGGCGTTAAACGTGTTAGGACTTCTTCTTACTCATAT 1526
```

```
384 heglyluAlaAspValGlnGluTrpAlaLysGluArgLysleuThrarg 400
|||||
1527 TTGGAGAGCAGATGCTCAAGATGGGCTTAAAGAGAAAACTCCTGCT 1576
401 Leu 401
|||
1577 TTA 1579
seq_name: gb_prl:AB020527
seq_documentation_block:
LOCUS AB020527 2626 bp mRNA PRI 14-MAY-1999
DEFINITION Homo sapiens mRNA for Na/P04 cotransporter homolog, complete cds.
ACCESSION AB020527
VERSION AB020527.1 GI:4587206
KEYWORDS Na/P04 cotransporter homolog.
SOURCE Homo sapiens tissue_l1b:small intestine cDNA to mRNA,
clone:kaia2138.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (sites)
AUTHORS Shibui,A., Tsunoda,T., Seki,N., Suzuki,Y., Sugane,K. and Sugano,S.
TITLE Isolation and chromosomal mapping of a novel human gene showing
homology to Na/P04 cotransporter
J. Hum. Genet. 44 (3), 190-192 (1999)
JOURNAL
MEDLINE 99253143
REFERENCE 2 (bases 1 to 2626)
AUTHORS Sugano,S. and Shibui,A.
TITLE Direct Submission
JOURNAL Submitted (27-NOV-1998) to the DDBJ/EMBL/GenBank databases. Sumio
Sugano, Institute of Medical Science the University of Tokyo,
Department of Virology; 4-6-1 Shirokanedai, Minato-ku, Tokyo
108-8639, Japan (E-mail: ssugano@ims.u-tokyo.ac.jp,
Tel:81-3-5449-5286, Fax:81-3-5449-5416)
location/Qualifiers
source
1..2626
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="kaia2138"
/map="6p22.3"
/tissue_l1b="small intestine"
/note="vector pME18S"
109..1602
/codon_start=1
/product="Na/P04 cotransporter homolog"
/protein_id="BAA76663.1"
/db_xref="GI:4587207"
/translation="MSTGDPVRAIVYGDIDSSDGNLVNAOECNRKGFCSVRHGLALITQ
LCNFSYTOQMLSLAIAPMNVNTAPSPQSNASTERSPSDSGYNMETLKEFKAMAPA
YMSPEIOGIIILSLNYSFLAIPSGYAGIFGAKYVYAGAGLIFSSFTLPIPLAN
AGVALLIVIRIOGIAOVNVLNGOVSIVKNAPIELRSQLTITAGSGMSLSPYLLA
GSLICTOTMPVYVFYIFGICGACCPMLPPLYDDPVNHPFISAGKRRTIVCSLAOD
CSPGMSLPKRAIKSLPLMAILVSTFCERWLVTYMAVPTVYISSVLQNLDSGLS
ALPFWGICITIIIGLADFLSLRKRLITRTKRLTAIGVAFPSVYLALDPSVSSH
SMTMTFLVLSAISISECSGALVNFIDIAPIRYTGFLKGLQVFAHLAGVISPAGGF
ISDSEFGGRNVFLLSAVNIISGLVYLLIFGADYQDMAKKEITFTLL"
BASE COUNT 648 a 651 c 560 g 767 t
ORIGIN
```

```
alignment_scores:
Quality: 1023.50 Length: 465
Ratio: 3.270 Gaps: 2
Percent Similarity: 67.312 Percent Identity: 43.441
```

```
alignment_block:
US-09-391-958-1 x AB020527
Align seg 1/1 to: AB020527 from: 1 to: 2626
```

```

16 CysSerAlaArgTyrGlyIleAlaLeuValLeuHisPheCysAsnPheth 32
   |||.....|
205 TGTTCAGTCGACATGGGCTGGCCCTCATCTTGCAGCTCTGTATTTTTC 254
   |||.....|
32 rThlIleAlaIleAsnValIleMetAsnIleThiMetValAlaMetValA 49
   |||.....|
255 AATTTCACCCACCAAAATGAATGAGCTGAGCATTCGACATTCACCTATG 304
   |||.....|
49 snSerThrSerProGlnSerGlnLeuAsnAspSerSerGlnValLeuPro 65
   |||.....|
305 ACAACAGACGCCCCACCTAGCCAGCCCATGGCTCCACAGAAAGGCCCTCC 354
   |||.....|
66 ValAspSerPheGlyIleGlyLeuSerLysAlaProLysSerLeuProAla 82
   |||.....|
355 ACTGACTCCGAGGGCTACTGGAATGAATCTTAAGAATTTTAAGCAAT 404
   |||.....|
82 sSer.....|
405 GGGCCCTGCATATGACTGAGTCTGTAATTCAGGAATCATCTCCAGCT 454
   |||.....|
83 .....|
455 CCCTCACTATAGGCTCATCTTGGCTCCAAATCCCACTGGCTATGGCT 504
   |||.....|
83 .....|
505 GGAATATTTGGAGCCAGATGATGGTGGTGGCTGGCTTTATTTTCCTG 554
   |||.....|
83 .....|
555 ATTCCGAGCCCTCTTCATTCCTACTGGCAGCTAATGGGGAGTGCCCTGC 604
   |||.....|
84 .....|
605 TCATTGTCTCCGAAATGTACAGCATAGCCAGGTTATGTAATTAAC 654
   |||.....|
88 GlyGlnPheAlaIleTyrGlnArgTyrPglyProProGlnIleArgSerAr 104
   |||.....|
655 GGCAGATTCATTCATTTGGTGCAATGGGCTCCCACTGGAAAGGAGTCA 704
   |||.....|
104 gLeuCysSerIleAlaLeuSerGlyMetLeuGlyCysPheThiAlaI 121
   |||.....|
705 ACTCACCACCATGTGTGATCAGGATCAGGATCGGGGGTCTCTATGTTTC 754
   |||.....|
121 leuLeuIleGlyIlePheIleSerGlnThiLeuGlyTyrProPheValAl 137
   |||.....|
755 TACTTGCTGGTGGTCTCTCTGCCAGACCATAGATGGCTTACGCTTC 804
   |||.....|
138 TyrIlePheGlyIleValGlyCysValCysCysLeuLeuThiPheValA 154
   |||.....|
805 TATATCTTTGGAGGAAATGGCTGTCTGTTTCTCTCTGCTGTTCTCT 854
   |||.....|
154 IleTyrAspAspProValSerTyrProTyrIleSerThrSerGluLysG 171
   |||.....|
855 CATTATATGATGATCCGTGAATCATCCCTTATCATGCTGTGGAGAAGA 904
   |||.....|
171 IuTyrIleIleSerSerLeuLysGlnIleValGlySerSerLysGlnPro 187
   |||.....|
905 GATACATTTGTGTTTCATTTGCTGCTCAGCAGACTGTTCACACAGCTG 954
   |||.....|
188 leuProIleLysAlaMetLeuArgSerLeuProIleTyrSerIleCysAr 204
   |||.....|
955 CTTCACATTAAGGCTATGATCAATCCCTTACACACCTGGCCATTTTACT 1004
   |||.....|
204 uGlyLysPheSerHisGlnTyrIleValSerThiMetValValTyrIleP 221
   |||.....|
1005 CTCTATTTCTGTGAATCTGCTTTTATACATTAATGCGGTACACAC 1054
   |||.....|
221 roThrTyrIleSerSerValTyrHisValAsnIleArgAspAsnGlyLeu 237
   |||.....|
1055 CAAGCTACATCAGCTCGGTACTTCAAGCCCACTGAGAGATAGTGGATC 1104
   |||.....|
238 leuSerAlaLeuProPheIleValAlaTyrValIleGlyMetValGly 254

```

```

1105 |||.....|
CTGTCTCCCTTGGCTTGTGGATGATGATCATTCCTTGGAGG 1154
254 YTrLeuAlaAspPheLeuThiLys...PheArgLeuIleThy 270
   |||.....|
1155 TCTACTGGCAGACTTTCTCTCCAAATAATCCCTAGACTATACCA 1204
   |||.....|
270 aAlArgLysIleAlaThiLeuGlySerLeuProSerSerAlaLeuIle 286
   |||.....|
1205 TCAGGAATCTTCACATCCATTTGGGGTCTCTCCATCCGGAATC 1254
   |||.....|
287 ValSerLeuProTyrIleAsnSerGlyTyrIleThiAlaThiAlaLeu 303
   |||.....|
1255 GTGTCCCTGGCTGGGTCAGATCCAGCCAGCATGACATGACCTT 1304
   |||.....|
303 urThiLeuSerCysGlyLeuSerThiLeuCysGlnSerGlyIleTyrIle 320
   |||.....|
1305 GGTGCTGTCTTCCCATCGACGCTTCTGTGAATCAGAGCCCTGT 1354
   |||.....|
320 snValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAlaSer 336
   |||.....|
1355 ACTTCTTGATATGCTCCTCGGTACACTGGCTTCTCAAGACTAT 1404
   |||.....|
337 ArgGlyPheSerSerIleAlaProValIleValProThiValSerGly 353
   |||.....|
1405 CAAGTCTTGACACATAGCTGAGCCATCTCTCTACTGCTGGAT 1454
   |||.....|
353 eLeuLeuSerGlnAspProGlnPheGlyTyrPArgAsnValPheLeu 370
   |||.....|
1455 TTTCATCAGTCAGGATTCAGAGTGTGGTGGAAATGCTCTTCT 1504
   |||.....|
370 eupHeAlaValAsnLeuGlyLeuLeuPheTyrIleIlePheGly 386
   |||.....|
1505 CAGCTCTGTAAATATCGGGCTGGTTTCTACCTCATCTTGGCCA 1554
   |||.....|
387 AlaAspValGlnIleTyrAlaLysGlnArgLysLeuArgLeu 401
   |||.....|
1555 GCAGATGTCAGAGACTGGGCTAAGAGCAGACATTCACCACCTC 1599
   |||.....|
seq_name: gb_pr2:HSNAPI1
seq_documentation_block:
LOCUS HSNAPI1 1549 bp mRNA PRI 25-JAN-1994
DEFINITION H.sapiens mRNA for sodium-phosphate transport system 1.
ACCESSION X71355
VERSION X71355.1 GI:450531
KEYWORDS brush border membrane; NPT1 gene; renal sodium-dependent phosphate transporter; transporter.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 1549)
AUTHORS Chong,S.S., Kristjansson,K., Zoghbi,H.Y. and Hughes,M.R.
TITLE Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3-p23
JOURNAL Genomics 18 (2), 355-359 (1993)
MEDLINE 94117004
REFERENCE 2 (bases 1 to 1549)
AUTHORS Chong,S.S.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1993) S.S. Chong, Baylor College of Medicine, Institute for Molecular Genetics, One Baylor Plaza, Houston, Texas 77030, USA
COMMENT Related sequence: M76466.
FEATURES
FEATURES
1..1549
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/dev_stage="adult"
/clone_lib="lambdaMAX kidney-cortex cDNA (Clontech)"

```

```
gene      /map="6p21.3-p23"
CDS       13..1416
          /gene="NP1"
          13..1416
          /gene="NP1"
          /codon_start=1
          /product="Sodium-Phosphate Transport System 1"
          /protein_id="CAA50490.1"
          /db_xref="GI:450532"
          /db_xref="SWISS-PROT:Q14916"
          /translation="MMDNRLEPPKKVPKPCFGRGLSLVHCCNVIIAORACLMITM
          VVWVSTDPHGLPNTSTKLLDINKPMYNMSPDIOGIIISTSYGIIIOVPGVFS
          GIVSKMIGFALCISVLSLIPPAAGIVAAVYCAVGAAGIVATAQOFIYVK
          WAPLERGLTMSRSGRLGPIFYLVTVGYCESLGMVPTIYAGACACLMFV
          LFYDDPKRPHCISISEKEYITSSLYOYSSRSOLPIAIIKSLFVMAISGFTFV
          SHNMTLYTPMFINSMHVNIKENGLSSLPYLFAMWIGNLGOLDFELRNLSVI
          AVRKLFYTAGFLPAIFGVCLPYSIFYIYFILAGATSCFGLGVFINGLDIAP
          RVFGIKACSTLTGMIGSLIYTLGLIKODPESAMFKTFLMAINVTGLIFLYLI
          ATAETODAKKEKHRL"
BASE COUNT 381 a 356 c 345 g 467 t
ORIGIN
alignment_scores:
  Quality: 964.00      Length: 467
  Ratio: 3.080        Gaps: 3
  Percent Similarity: 67.024  Percent Identity: 41.328
alignment_block:
US-09-391-958-1 x HSNAP11 ..
Align seg 1/1 to: HSNAP11 from: 1 to: 1549
1 MetGlnValaspGluThrLeuIleProAlaGlyValaProSerLeuCysSe 17
  |||||:||||: ||| |||||:||||:||||:||||:
13 ATCCAATGATGATAACCGGTGCTCCCAAAAAGTTCCAGGTTCTGTTC 62
  |||||:||||:||||:||||:||||:||||:
17 TAlaArgTyrGlyIleAlaLeuValLeuHisPheCysAsnPheThrIrt 34
  | |||||:||||:||||:||||:||||:
63 CTTTCCTATGATGATGCTTCTTCTTGCACTGTGTAAATGTTATAATAA 112
  |||||:||||:||||:||||:||||:
34 LeuIaGlnAsnValIleMetAsnIlePheMetValaIleMetValaSer 50
  |||||:||||:||||:||||:||||:||||:||||:
113 CACGACAGCGTGGCTGCCTCAACCTCATGATGATGATGATGATGATG 162
  |||||:||||:||||:||||:||||:||||:
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeu..... 64
  |||||:||||:||||:||||:||||:
163 ACAGATCCACATGATGTTGCCACACACCTCCACAAAGAACTCCTGGATA 212
  |||||:||||:||||:||||:||||:
64 ..... 64
213 TATAAAGAACCTATGATTAATTGAGACCCAGATATCCAGGAAATCATCT 262
  .....
65 ..... ProAlaAspSer 68
  .....
263 TGAATCCACCTCCTATGATGATCATCATCCAAAGTTCTGTGGATAC 312
  |||||
69 PheGlyGlyLeuSerIleValaProLysSerLeu..... 79
  |||||:||||:||||:||||:||||:
313 TTCTCTGCAATATATATCTACAAAGAAATGATGCTTGCATTATATCCT 362
  .....
79 ..... 79
363 CACGCTCTGTATTAAGCTGTCTCATCCACAGCAGCTGAAATGAGTAG 412
  .....
80 ..... ProAlaLysSerSerIle 85
  .....
413 CTGGGCGCTGTATCTGACAGCATTCAGGAGACGCCAGGGGATAGTT 462
  .....
86 LeuGlyGlyIlePheAlaIleTyrGluArgTyrPylProGlnGluArg 102
  ::|||:||||:||||:||||:||||:||||:
463 GCAACAGCCAGTTTGAATATATATGTCAAATAGGCGCTCCCTCGAAGC 512
```

```
102 gSerArgLeuCysSerIleAlaLeuSerGlyMetLeuGlyCysPhe 119
  |||||:||||:||||:||||:||||:||||:
513 AGCCGACTTACTTCTATGATACATCAGGGTTTCTGGACCTTTA 562
  .....
119 hAlaIleLeuIleGlyIlePheIleSerGlnThrLeuGlyTyrProPhe 135
  |||||:||||:||||:||||:||||:||||:
563 TTGTCCTACTTGTGACATGAGATTATCTGTGAATCTCGGGCTGCCATG 612
  .....
136 ValPheTyrIlePheGlyGlyValGlyCysValCysLeuLeuTyrPhe 152
  |||||:||||:||||:||||:||||:||||:
613 GTCTCTCATATTTTGGTGTGCTGAGCTGCTCATGCTCTCTCTGTT 662
  .....
152 eValValIleTyrAspAspProValSerTyrProTyrIleSerThrSerg 169
  |||||:||||:||||:||||:||||:||||:
663 CGTTCTGTTATGATATACCCCAAGACCCCAATGTAAGCATCAGT 712
  .....
169 LuYsgLutYrIleIleSerSerLeuLysGlnValGlySerSerLys 185
  |||||:||||:||||:||||:||||:||||:
713 AAAAGAAATACATCATCTCCTCCCTGCTCCAGCAGTCAGTCAAGTAGA 762
  .....
186 GlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTyrSerI 202
  |||||:||||:||||:||||:||||:||||:
763 CAATCTCTGCCATCAAGCTATCAAGCTTAAAGCTTCCAGTCTGGCTAT 812
  .....
202 eCysLeuGlyCysPheSerHisGlnTyrPheValSerThrMetValTyr 219
  |||||:||||:||||:||||:||||:||||:
813 TTCCATTGATGATGATGATGATGATGATGATGATGATGATGATGAT 862
  .....
219 YrIleProThrTyrIleSerSerValTyrHisValaIleArgAspAsn 235
  |||||:||||:||||:||||:||||:||||:
863 ACATCCCAATGTTATATCACTCCATGCTTCATGTTATATATAAGAAAT 912
  .....
236 GlyLeuLeuSerAlaLeuProPheIleValAlaTyrValIleGlyMetVa 252
  |||||:||||:||||:||||:||||:||||:
913 GGGTCTTCTGCTTCCCTCATTTGTTGCTGCGATCTGTGTAACCT 962
  .....
252 IGLGlyTyrLeuAlaAspPheLeuThrIleLys... PheArgLeu 268
  |||||:||||:||||:||||:||||:||||:
963 AGCAGGTACATGATACATCTTCTGACCAAGAAATTTCTCAGCGTAA 1012
  .....
268 IeThrValArgLysIleAlaThrIleLeuGlySerIleProSerSerAla 284
  |||||:||||:||||:||||:||||:||||:
1013 TTGCTGTCGCGAATCTTTCACACACAGCATTTCTCTCCGCAATC 1062
  .....
285 LeuIleValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAl 301
  |||||:||||:||||:||||:||||:||||:
1063 TTGCTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1112
  .....
301 aLeuLeuThrLeuSerGlyLeuSerThrLeuGlyGlnSerGlyIleT 318
  |||||:||||:||||:||||:||||:||||:
1113 TTTCCTATATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1162
  .....
318 YrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGly 334
  |||||:||||:||||:||||:||||:||||:
1163 TTTAATAATGCTTGTGATATCTCTCCACAGATATTTGGATTTATTAAGCA 1212
  .....
335 AlaSerArgGlyPheSerSerIleAlaProValIleValProThrValSe 351
  |||||:||||:||||:||||:||||:||||:
1213 TGTTCACACTTAACTGGAATGATGAGAGACTAATTTCTTCCACTTTGAC 1262
  .....
351 rGlyPheLeuLeuSerGlnAspProGluPheGlyTyrPArgAsnValPheP 366
  |||||:||||:||||:||||:||||:||||:
1263 TGGATTTATCTCTTAAGCAGGATCCGGAATCCGCTGTGTTAAACCTTCA 1312
  .....
368 hLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuIlePhe 384
  |||||:||||:||||:||||:||||:||||:
1313 TCTCGATGCGACCATTAATGATGATGATGATGATGATGATGATGATGAT 1362
  .....
385 GlyIleAlaAspValGlnGluTyrPalaLysGluArgLysLeuThrArgLe 401
  ::|||:||||:||||:||||:||||:||||:
1363 GCTACACAGCAATATTCAGAGCTGGGCTTAAAGAAAACAAACACACAGCT 1412
  .....
401 u 401
```


REFERENCE	1 (bases 1 to 1794)
AUTHORS	Miyamoto, R.-I.
TITLE	Direct Submission
JOURNAL	Submitted (17-FEB-1994) to the DDBJ/EMBL/GenBank databases.
REFERENCE	Ken-ichi Miyamoto, School of Medicine, University of Tokushima, Department of Nutrition; 3 Kuramoto-cho, Tokushima, Tokushima 770, Japan (Tel:0886-31-3111(ex.2527), Fax:0886-33-7094)
AUTHORS	2 (bases 1 to 1794)
TITLE	Miyamoto, R., Tatsumi, S., Sonoda, T., Yamamoto, H., Minami, H., Takekoshi, Y., and Takeda, E.
JOURNAL	Cloning and functional expression of a Na(+)-dependent phosphate co-transporter from human kidney: cDNA cloning and functional expression
MEDLINE	Biochem. J. 305 (Pt 1), 81-85 (1995)
FEATURES	95126933
source	Location/Qualifiers
gene	1..1794
CDS	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/dev_stage="adult"
	/sex="male"
	/tissue_type="kidney cortex"
	4..1401
	/gene="Na/Pi-4"
	4..1401
	/gene="Na/Pi-4"
	/codon_start=1
	/product="renal Na+-dependent phosphate cotransporter"
	/protein_id="BA05888.1"
	/db_xref="gi:639842"
	/translation="MDNRILPKRVPCSPFPGSLFVHCNVIITTAQMLCNTMNVV MNSDPLHGLPNTSTKRLDNIRKPMKNSMSPDVGITLSTSGVILITQVPVGFSGI YSTRKMTGFRALCTSSVSLPIPAAGIGVAMVVCAVGAAGIVATAQELIVKNA PLERGRLLTSMSTGSLGPEFLVLTGVIGESLGMWPFYIFGACCAVACOLWFLV YDPEKDPHCISIGEKVETSLVQVSSRSQSLPKALKLSLPWALISIGSEPFMSH NIMLTPMEINSMIAHNKENGFLSLPYLFANIGNLGASDPLFTNRILSVIAW RKLTPTAGSLFLPAIFGVCLPISTFPIYVFLILACATGSCIGGVPIINGDIAPRY EGFLKAGSTLTMGGILASITNGLILKQPEBSAMFTFLMAINTVGLIFLIVAT AEIODMAKEKHTRL"
BASE COUNT	460 a 398 c 385 g 551 t
ORIGIN	
alignment_scores:	Quality: 944.00 Length: 466
	Ratio: 3.045 Gaps: 3
Percent Similarity:	66.524 Percent Identity: 40.773
alignment_block:	
US-09-391-958-1 x HUMAP14 ..	
Align seg 1/1 to: HUMAP14 from: 1 to: 1794	
2	GINVAIAPGLUthLeuIIePrOArGLySvAlProSerLeuCySeRal 18
1	CAAAATGATTAACCGGTGCTGCCCAAAAAAGTTCACAGGTTCTGTCCT 50
18	aArGTyGcLYIleAlaLeuValLeuHISphecYasAnPheThrTrIleA 35
51	CCGCTATGATTTGCTTTCCTGTGCACTGTGTATATGTTAATAAACAAG 100
35	IaGInAsnValIleMeLcAsnIleThrMetValAlaMetValAsnSerThr 51
101	CACAGCTATGTGCCTGAACCTCAACAATGGTGTGATGATGTGAATACACA 150
52	SePrOgInSerGInLeuAsnAspSerSerGInValLeu..... 64
151	GATCCACATGGTTGCCCAACACCTCCCAAGAAGCTCTCGATATATAT 200
64 64
201	AAAGACCTATGATATATGGAGCCCAAGATCTCCACGGAATCATCTTGA 250

```

65 .....ProValaspSerPhe 69
251 GTTCAACCTCTATGGTGCATCATCATCAAGTCTCTGTGATATCTC 300
70 GYGLYLeuSerLysAlaProLysSerLeu.....79
301 TCTGGATATATTCTACAAAGAAATGATTGCTTGCATTATGCTCAG 350
79 .....79
351 CTCTGTGTTAAGCTGCTCATCCACAGACAGCTGCAATTGAGTACCTT 400
80 .....ProAlaLysSerSerLeu 86
401 GGGTCGTGTATGTCAGCAGTTCAGGAGACCCAGGGGAGATTGCA 450
87 GYGLYGLNpheAlaIleTTPGLuATTPGLYProProGLNGLuATPse 103
451 ACAGCCAGATTGAAATATATGTCAAATGGGCTCTCCCTGGAGCAGG 500
103 rArgLeuCySerLLeaLaleuSerGlyMetLeuLeuGlyCyPheThra 120
501 CCACCTACTTCTATGATGATCAGATGAGGCTTTGCTGGGACCTTATTT 550
120 lAlleuLleGLYGLYpHeLleSerGLuThrLeuGLYTPProPheVal 136
551 TCTCTACTTGACTGAGATATCTGTGATCTCTGGCTGGCCATGCTC 600
137 PheTYrLlePheGLYGLYGLYGLYGLYGLYGLYGLYGLYGLYGLY 153
601 TCTATATTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 650
153 lValLleTYAspAspProValSerTYrProTPPLeSerThSerLul 170
651 TCGTTTATGATGACCCCAAGACCAACCAATGATATAGATGATGATA 700
170 ySGLuTYrLleLleSerSerLeuLysGLNGLNValGLYSerSerLysGLN 186
701 AGGAATATACATCATCTCCCTGCTGCAGCAGCATTCACAGTACAA 750
187 ProLeuProLleLysAlaMetLeuArgSerLeuProLleTPSerLecy 203
751 TCTCTGCTATCAAGCTATATTAAGTCCCTTCCACTGCGCTATTTC 800
203 sLeuGLYCyPheSerHisGLNThrLeuValSerThMetValTYrI 220
801 CATGGATGTTTACGTTTCTGTCACATATCATCATGACATATACA 850
220 lProThTYrLleSerSerValTYrHisValAsnLleArgAspArgLy 236
851 CTCGAATGTTTATCACTCCATGCTTCTATATATATAAGATGCG 900
237 LeuLeuSerAlaLeuProPheLleValAlaTYrValLleGLYMetValGL 253
901 TTTCTGTCTCCCTCCCTATTTGTTGCTGATCTGTGTAACTCAGC 950
253 yGLYTYrLeuAlaAspPheLeuLeuThrLysLys...PheArgLeuLleT 269
951 AGCTCACTATATCAAGACTTCTCTGACAGCAAGAAATATTCAGCTAATTG 1000
269 hValAlaGLYsLleAlaThrLleLeuGLYSerLeuProSerSerAlaLeu 285
1001 CTGTCCGAAACTCTTCACAGCAGCAGGATTTCTCTTCTGCAATCTTT 1050
286 lLeValSerLeuProTYrLeuAsnSerGLYTYrLleThraLleThraLle 302
1051 GGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1100
302 uLeuThrLeuSerCyGLYLeuSerThrLeuCyGLNGLNGLYLeTYrI 319
1101 CTTAATATCTTGTGTGTCACACAGCAGCTTTGTTGTTGAGATGTTTA 1150
319 lAsnValLeuAspLleAlaProArgTYrSerSerPheLeuMetGLYAla 335

```

```

1151 TAAATGGCTTGATATGCTCCAGATATTTTGGATTATTAAGCATGT 1200
336 SerArgLysPheSerSerLleAlaProValLleValProThraValSerL 352
1201 TCAACTTAAGTGAATGATAGAGGACTATATCTTCCACTTGTGACTGG 1250
352 yPheLeuLeuSerGLNAspProGLuPheGLYTPArgAsnValPhePheL 369
1251 ATTGATCTTAAAGCAGATCCGAGATCCGCTGTTTAAACCTTCATCC 1300
369 euleuPheAlaValAsnLeuLeuGLYLeuLeuPheTYrLeuLlePheGLY 385
1301 TGATGGCAGCATTAATGATGACTGCTTAATTTCTACCTATATGTTGCT 1350
386 GluAlaAspValGLNGLUTPAlaLysGLUArgLysLeuThrArgLeu 401
1351 ACAGCAGAAATTCAGAGCTGAGCTAAAGAAACACACACACGCTC 1398

```

seq_name: gb_com:RABRCCTX

seq_documentation_block:

LOCUS RABRCCTX 1855 bp mRNA MAM 27-APR-1993

DEFINITION Rabbit renal cortical Na/P-1-cotransporter, complete cds.

ACCESSION M76466

VERSION M76466.1 GI:165689

KEYWORDS membrane transport protein; renal cortical Na/P-1-cotransporter; transmembrane protein.

SOURCE Oryctolagus cuniculus kidney cortex cDNA to mRNA.

ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

REFERENCE 1 (bases 1 to 1855)

AUTHORS Werner,A., Moore,M.L., Mantei,N., Biber,J., Semenza,G. and Murer,H.

TITLE Cloning and expression of cDNA for a Na/Pi cotransport systems of kidney cortex

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 88, 9608-9612 (1991)

MEDLINE 92052140

FEATURES

source

Location/Qualifiers

1..1855

/organism="Oryctolagus cuniculus"

/db_xref="taxon:9986"

/tissue_type="Kidney cortex"

116..1513

/note="expression in Xenopus laevis oocytes"

/codon_start=1

/product="renal cortical Na/P-1-cotransporter"

/protein_id="AAA31461.1"

/db_xref="GI:165690"

/translation="MDNOPSRKGRFCFSRYVALFMHFCNIVIIQAMCSTINVA

AVNNNTNHSPTSAERLDNKNPYNNSPVQGITFSSITVGAFLQIPVGIISGI

YSIKRLGFALFLSLVSIPIQAAVGETWIVICRVVQIGTVTTAHOHIVKWA

PLERGLRSLMSGLFLGPFLVLTAVTGIIISLGMVNFYIFGACGACVLIEMVLY

YDDPKDHPCSVLSHEKEYITRSSLIQOGSSVRSOLPIKAMKSLPIKAMISRCFAYIMTY

SLRIYTPILINSMLHVDIRENGISLPIELAMVICAGHTADPFLMRNLSTAI

RKLFTRIGLLPIVFSMCLTYLSSGFTSTITLLIANNSSSCDGLGALLNALDLPRI

YVFIRGVTLIGMTGMSSTVAGFLSODPSSSWPKIFLLMSIINVISVIFLLFAK

ABIQMAKKEQKQTRL"

ABIDOMAKKEQKQTRL"

1836..1841

polyA_signal

BASE COUNT 459 a 459 c 393 g 544 t

ORIGIN

alignment_scores:

Quality: 937.00 Length: 466

Ratio: 3.023 Gaps: 3

Percent Similarity: 66.524 Percent Identity: 39.485

alignment_block:

US-09-391-958-1 x RABRCCTX ..

Align seg 1/1 to: RABRCCTX from: 1 to: 1855


```

||||| |||:||||| |||:|||||:
440 AGCAAGATCCAGATTCTGTCATTACGCTATGGCGTGGCTTATCAT 489
26 uhlsphecyasnphehrthrilleaglnasnvalillemtasnlier 43
   ||||| ||||| |||||:|||||:|||||:
490 GCACTTCTCAACTCCACATGATTAACGCGGTGTGATGTCAAGCATTC 539
43 hrmetvalametvalasnserthrserproglinserglnleuasnasp 59
   ||||| ||||| |||||:|||||:|||||:
540 GATTCATCCGCATGTGTACACACCACTCAGACGAGGTCTATTAATGCC 589
60 SerSerGluValLeuProValAspSerPheGlyLeuSerLysAlaPr 76
   ||||| |||:|||||:|||||: |||
590 TCCACTGAGGGGCGCTGTGCAGATGCTTCAATACATCCAGCATATTCAT 639
76 olysSerLeuProAlaLysSerSerlle..... 85
   |||||:|||||:|||||:
640 CAAGGAATTTGATACAAAGGCTGTGTATCAATGAGGCCAGAAATC 689
85 ..... 85
690 AGGATATCATCTTACCTCCATCACTATGGATAACTGACTGTGATC 739
86 ..... leugly.. 87
740 CCAAGTGTGATATTAGCAGGATATTGGAGCAAAAAAATGCTGTGTC 789
87 ..... 87
790 TGGTTGCTGATCTTCCCTCTCAACCTTATTACACACATGGCTGCT 839
87 ..... 87
840 ACTTCGAGATTTGGTCATCATGTTGGACAGTCGAGGCCATGGCC 889
88 ..... glylnphealalletrpgluartrpjlpr 98
890 CAGGAATGCATGACAGGTCAGATTACTATTGGCAAAAGTGGCTGC 939
98 oProGlnGluArgSerArgLeuGlySerlleAlaLeuSerGlyMetLeu 115
   ||||| |||||:|||||: |||||
940 TCCACTTGAAGCAAGCAAGCTCACACCATGTCAGATGAGGTGCACAT 989
115 euGlyCysPheThrAlaIleLeuIleGlyLysPheIleSerGluThrLeu 131
   ||||| |||:|||||:|||||:|||||:
990 TTGGATCTTCATCATCCTCTGTGTGGGGGACTAATCTCAAGGCTTG 1039
132 GlyTrpPheValPheTyrIlePheGlyGlyValGlyCysValCysCy 148
   ||||| |||||:|||||:|||||:|||||:
1040 AGCTGGCCTTTATCTTCTACATCTTGTGACACTGCTGTGTCTGTG 1089
148 sleuLeuTrpPheValValIleTyrAspAspProValSerTyrProTrp 165
   ||||| |||||:|||||:|||||:|||||:
1090 TCTCTATGCTTACAGTCACTGATTATGATGACCCATCATCACCCTG 1139
165 leSerThrSerGluLysGluTyrIleIleSerSerLeuLysGlnGlnVal 181
   ||||| |||||:|||||:|||||:|||||:
1140 TAAGTGTAGGAAAGAGACACATCTGCTCTCAGCTCAACACACCC 1189
182 GlySerSerLysGlnProLeuProIleLysAlaMetLeuArgSerLeuPr 198
   ||||| |||:|||||:|||||:|||||:
1190 AGTTCTCCTGAGAGAGCTGCCCATTAAGGCGATGTCAATGCTCTAC 1239
198 oIleTrpSerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerT 215
   ||||| ||||| ||||| |||||
1240 ACTTTGGCCATTTTCTCGGTTTTTTCACCCATTTCTCGTTGTGACCA 1289
215 hrMetValValTyrIleProThrTyrIleSerSerValTyrHisValAsn 231
   ||||| |||||:|||||:|||||:|||||:
1290 TCATCTCAATACCTACCAAGATATATGACTGTCTCCATGTATAC 1339
232 IleArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpVa 248
   ||||| |||||:|||||:|||||:|||||:

```

```

1340 ATCAGAGATAGTGGAGTCTGTCTCCCTCCCTTTATTCCTGTCGCAAG 1389
248 lIleGlyMetValGlyLysTyrIleuAlaAspPheLeuThrLysLys. 264
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1390 CTGTACAAATTTTAGAGGTGACGTGGCAGATTTCTCTTTCACGAAATC 1439
265 .PheArgLeuIleThrValArgLysIleAlaThrIleLeuGlySerLeu 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1440 TTCTCATATTGATCATCTGTGCGAAAGCTCTTTCATCTTGA..... 1482
281 ProSerSerAlaLeuIleValSerLeuProTyrLeuAsnSerGlyTyrIl 297
1482 ..... 1482
297 eThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuCysG 314
1482 ..... 1482
314 lnsrGlyIleTyrIleAsnValLeuAspIleAlaProArgTyrSerSer 330
   ||||| |||||
1483 ..... TATGCAAT 1491
331 PheLeuMetGlyAlaSerArgGlyPheSerSerIleAlaProValIleVa 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1492 TTCCTCATGGGAATCTCAAGGGATTTGGCTCTCATCGCAGAAATCATCTC 1541
347 lProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpA 364
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1542 TTCACATGCCATCGATTCCTCATCATCAGATTTGAGTGTGTGA 1591
364 rGAsnValPhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuPhe 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1592 GGAATGTCTTTTCTCTCTGCTGCTGACATCAACATGTTGGCTGTCTTT 1641
381 TyrLeuIlePheGlyGluAlaAspValGlnGluTrpAlaLysGluArgLy 397
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1642 TACCTACGTTTGGACACAGACGAACTTCAAGACTGTGGCCAAAGAGAGAC 1691
397 sleuThrArgLeu 401
   ||||| ||||| |||||
1692 CCTTACCCGCTC 1704
seq_name: gb_p13:HSU90544
seq_documentation_block: 2281 bp mRNA PRI 02-MAY-1997
LOCUS HSU90544 Human sodium phosphate transporter (NPT3) mRNA, complete cds.
DEFINITION Human sodium phosphate transporter (NPT3) mRNA, complete cds.
ACCESSION U90544
VERSION U90544.1 GI:2062689
KEYWORDS
SOURCE
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euteria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 2281)
Ruddy,D.A., Krommal,G.S., Lee,V.K., Minter,G.A., Quintana,L.,
Domingo,R. Jr., Meyer,N.C., Basava,A., McLelland,E., Fullan,A.,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchinashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
A 1.1 megabase transcript map of the human hereditary
hemochromatosis locus
JOURNAL
REFERENCE 2 (bases 1 to 2281)
Ruddy,D.A., Krommal,G.S., Lee,V.K., Minter,G.A., Quintana,L.,
Domingo,R. Jr., Meyer,N.C., Basava,A., McLelland,E., Fullan,A.,
Mapa,F.A., Moore,T., Thomas,W., Loeb,D.B., Harmon,C.,
Tsuchinashi,Z., Wolff,R.K., Schatzman,R.C. and Feder,J.N.
DIRECT SUBMISSION
JOURNAL Submitted (25-FEB-1997) Sequencing, Mercator Genetics, 4040
Campbell Avenue, Menlo Park, CA 94025, USA
FEATURES
location/Ouallifiers
1..2281
/organism="Homo sapiens"

```

```

alignment_scores:
  Quality: 932.00      length: 4771
  Ratio: 3.305         Gaps: 4
  Percent Similarity: 59.673      Percent Identity: 41.8208
alignment_block:
  US-09-391-958-1 x HS090544 ..
Align seg 1/1 to: HS090544 from: 1 to: 2281

```

98 oProGlnGlnArgSerArgLeuGlyCysSerTLeAlaLeuSerGlyMetLeuL 115
||||| ||||||| ::::: ||||| |||||
940 TCACATTGAAAGAAAGCAAGCTCACACACCATTTGGACGATGAGGTACGAT 989
115 euGlyCysPheThrAlaIleLeuIleGlyAlaPheIleSerGlnThrLeu 131
::: ||| ||| ::::: ||||| ::::: |||||
990 TTGGATCCTTCATCATCTCTGTCGTGGGGGAGCTAATTCACAGGCTTG 1039
132 GlyThrProPheValPheTyrIlePheGlyGlyValCysValCysCy 148
::: ||||| ::::: ||||| ::::: |||||
1040 AGCTGGCCTTTTACCTTCACATCTTTGGTAGACAGCTGCTGTCGCG 1089
148 sLeuLeuThrPheValValIleTyrAspAspProValSerTyrProTpi 165
TCTCCTATGCTTCACAGTATTTATATATACCCATGACATCACCCGTGCA 11399
165 lSerThrSerGlyLeuGlyGlyTyrIleIleSerSerLeuGlyGlnVal 181
||||| ::::: ||||| ::::: |||||
1140 TTAATCTTAGGAAAAAGAGACATCTCTGCTCCACTGGCTACAGGCC 11899
182 GlySerSerLeuGlnProLeuProIleGlyAlaMetLeuArgSerLeuP 198
::: ||| ::::: ||||| ::::: |||||
1190 AGTTCCTCTGGACAGAGCTGCCCCATAAAGGAGATGGACATGGCTATCC 12399
198 oIleTyrSerIleCysLeuGlyCysPheSerHisGlnThrLeuValSer 215
||| ::::: ||||| ||||| |||||
1240 ACTTTGGGCATTTTCTCTGGGTTTTTTCAGGCATTTTGCGTTATGACACA 12899
215 hMetValValIlyThrIleProThrTyrIleSerSerValTyrHisValAsn 231
||||| ::::: ||||| ::::: |||||
1290 TCATCTCAACATATCCATACCAACAGTATATCAGTACCTCTCCATAGTTAAC 13399
232 lLeArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaIleTyrpva 248
||| ::::: ||||| ::::: |||||
1340 ATCAGAGATAGTGAAGTCTGTGCCCTCCGCTTTATTCCTGCTGCACAG 13899
248 lIleGlyMetValGlyGlyTyrLeuAlaAspPheLeuLeuThrLysLys 264
::: ||||| ::::: ||||| ::::: |||||
1390 CTGTACAAATTTTAGGAGGTCACACTGGCAGATTTCTTTGTCACAGAAATC 14399
265 .. PheArgLeuIleThrValAlaGlySisIleAlaIleIleGlySerLeu 280
||| ::::: ||||| ::::: |||||
1440 TTCTCAGATTGATACACTGTGGCAAGCTCTTTATCTCTTGA..... 1482
281 ProSerSerAlaLeuIleValSerLeuProTyrLeuAsnSerGlyTyrIl 297
||| ::::: ||||| ::::: |||||
1482 1482
297 eThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuGlyc 314
||| ::::: ||||| ::::: |||||
1482 1482
314 lnsSerGlyIleTyrIleAsnValLeuAspIleAlaProArgTyrSerSer 330
||| ::::: ||||| ::::: |||||
1483 TATGCAGCT 1491
331 PheIleMetGlyAlaSerArgGlyPheSerSerIleAlaProValIleVal 347
||| ::::: ||||| ::::: |||||
1492 TTCTCCTATGCGCATCTCAAGGAGGATTTGGGCTATCCACAGATCATCTTC 1541
347 lProThrValSerGlyPheLeuLeuSerGlnAspProGlnPheGlyTyrpva 364
||| ::::: ||||| ::::: |||||
1542 TTCACCTGCGCATGATCTCCATCATCAGTCAAGATTTTGAATCTGGTTGGA 1591
364 rAsnValPhePheLeuLeuPheAlaValAsnLeuGlyLeuLeuPhe 380
||| ::::: ||||| ::::: |||||
1592 GGAATGCTTTTTCCTGCTGCTGCTGCTGCTCAACAATGTTGGCTGCGCTTT 1641
381 TyrLeuIlePheGlyGlnAlaAspValGlnGlyThrAlaLysLysLysArg 397
||| ::::: ||||| ::::: |||||
1642 TACCTCAGCTTTGGACAGCAAGCAACTTCAAGACTGGGCCCCAAAGAGGAGAC 1691
397 sLeuThrArgLeu 401

```
|||||
1692 CCTACCGCCTC 1704

seq_name: gb_hvg2:AL138726

seq_documentation_block:
LOCUS AL138726 140266 bp DNA HTG 15-FEB-2000
DEFINITION Homo sapiens chromosome 6 clone RP1-139G21, *** SEQUENCING IN
PROGRESS ***, 5 unordered pieces.
ACCESSION AL138726 GI:6982716
VERSION AL138726.2 GI:6982716
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 140266)
AUTHORS Sims,S.
TITLE Direct Submission
JOURNAL Submitted (15-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequests@sanger.ac.uk
On Feb 16, 2000 this sequence version replaced gi:6911804.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in
progress and the release of this data is based on the understanding
that the sequence may change as work continues. The sequence may
be contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Contig_ID: 00595 Length: 20679bp
Contig_ID: 00749 Length: 63415bp
Contig_ID: 00902 Length: 9637bp
Contig_ID: 01506 Length: 27403bp
Contig_ID: 01824 Length: 15932bp.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 20679: contig of 20679 bp in length
* 20680 21479: gap of 800 bp
* 21480 84894: contig of 63415 bp in length
* 84895 85694: gap of 800 bp
* 85695 95331: contig of 9637 bp in length
* 95332 96131: gap of 800 bp
* 96132 123534: contig of 27403 bp in length
* 123535 124334: gap of 800 bp
* 124335 140266: contig of 15932 bp in length.
FEATURES
source
1..140266
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="RP1-139G21"
/clone_1lb="RC1-1"
BASE COUNT 41912 a 26597 c 25588 g 42968 t 3201 others
ORIGIN

alignment_scores:
Quality: 876.50 Length: 503
Ratio: 4.039 Gaps: 7
Percent Similarity: 43.141 Percent Identity: 42.942

alignment_block:
us-09-391-958-1 x AL138726/rev ..
Align seg 1/1 to reverse of: AL138726 from: 1 to: 140266

141 G1g1yVal1gCysVal1gCysLeuLeuTrpHeVal1Val1IeTyraS 157
```

```
|||||
54076 GGAGTGTGTGGCTGTCTGCTGCTTCTCTGCTTGTGTGATTTATGA 54027
157 pAspProValSerTyrrProTrpIleSerThrSerGluLysGluTyrrIleI 174
|||||
54026 TGACCCCGTTTCCATCCATGATGATAGACACCTCAGAAAAAGATATCATCA 53977
174 leSerSerLeuLysGln..... 179
|||||
53976 TATCCTCCTGGAACAACAGGTACATATAAGACACCAACACTTTCTGA 53927
179 ..... 179
53926 GACCTGTATCTGATTTTCTTAGAGCCTTAATCTGATTTTAAAGGT 53877
179 ..... 179
53876 GACTAAGTGTGAGCATCTGCGATTACCGTTATGACCACTAAATTT 53827
180 .....GlnValGlySerSerLysGlnProLeuProIleLysAlaMetLeu 194
|||||
53826 GTTTTCAGGTCGGGCTCTTAAGCAGCCTTCCCATCAAAAGCTATGCTC 53777
195 ArgSerLeuProIleTrpSerIleCysLeuGlyCysPheSerHisGlnTr 211
|||||
53776 AGATCTTACCCATTTGGTCCATATGTTTAGCGTGTTCACCCATCAATG 53727
211 pLeuValSerThrMetValValTyrrIleProThrTyrrIleSerSerValT 228
53726 GTTAGTTAGCCATGGTGTATACATACCACTTACTACGCTGTGTGT 53677
228 YrHisValAsnIleArgAsp..... 234
|||||
53676 ACCATGTTAATCATCAGACGACGTAGTAGTTCCTTCCTTCCTTCTCC 53627
234 ..... 234
53626 TGCITGCATGGCTGACCAATTACTCTGCCCTCAATCATTCATCTGA 53577
234 ..... 234
53576 GAAATGATTTCTTATTACCAAAATATATCTAATTTTAATACATGCC 53527
234 ..... 234
53526 AGTAGCTTTCTGGGATGATTTAAGAGAGTTACTGATTTTCTTAAGAAA 53477
234 ..... 234
53476 GTATGATTTCAGAATGTCATTATTATTAGTAGTAATTTATACAAAAAGT 53427
235 .....A 235
53426 TGTGTCCTCACTGCTTTATTATAGTAATTACTCAATCTTTCTCTTA 53377
235 sn..GlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGlyM 251
:::|||||
53376 GAATGAGACTTCTATCTGCCCTTCTTATATGTTGCCGTGATAGGCA 53327
251 etValGlyGlyTyrrLeuAlaAspPheLeuLeuThrLysLysPheArgLeu 267
|||||
53326 TGGTGGAGGCTATCTGCGAGATTTCTCTTAACCAAAAAGTTTAGACTC 53277
268 IleThrValArgLysIleAlaThr..... 275
|||||
53276 ATCACTGTGAGAAATTTGCCACAATTTAGTAAAGAACAGCTTGAGG 53227
275 ..... 275
53226 GTCAGATTAATTTGCGTAGATACACAGCCAAAGATGCATCTCATGTT 53177
276 .....Ile..LeuGlySerLeuProSerSerAlaLeuI 286
|||||
```

seq_name: gb_pr3:HSA387747	seq_documentation_block:	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE
53176	TAATGATGTTCTTCTTCAATTCCTAGGAAGAGTGTCCCTCTTCAGACATCA	26	levalSerleuProTyrleuasnSerGlyTyrIleThAlaThAlaLeu	53126	TTGGTCTCTCTCCCTTACCTCAATTCGGCTATATCAGCAACTGCTTG	303	LeuThrIleuSerCysGlyIleuSerThreucysGlnSerGlyIleTyrIle	53076	CTGAGCGCTCTCTGGGAGTTAAGCACAATGTCAGTCAAGGATTAAT	319	easnValIleuaspIleala
52976	TTTCACATTTCTTTAAAGATTAGCTTTGACGTTTCTGGATCCCAAA	325	53026	CAATGCTTAATATATGCTCCAAAGGTAGAGCTGTATCTCCACTATAGT	325	52976	TTTCACATTTCTTTAAAGATTAGCTTTGACGTTTCTGGATCCCAAA	325
52926	CTTAGAATATGATTAACCATAGGGAGAGAACACAGATAGACCTGTGGC	325	52876	CGTTTTTTTTCAGGTATGTTGAGGAAGGGATGATTTTGTCTCATCCCT	325	52926	CTTAGAATATGATTAACCATAGGGAGAGAACACAGATAGACCTGTGGC	325
52826	CTCATTTGAATATACATAGTGCAGATGCTTTTGTATATATTTGTCTAATTA	325	52776	TAATTAATATATATTTGMAATATATATGACTATATATCATATCAGTGA	325	52826	CTCATTTGAATATACATAGTGCAGATGCTTTTGTATATATTTGTCTAATTA	325
52726	TCAGGGCTGAAGGATCTGTCMAACATCTCGGATCTAGACCGGACATTT	326Pro.ArgTyrSerSerPheIleuMetGlyAlaSerArgIle	52676	CTGTGACCCACACAGGTATTCACAGTTTCTCATGGAGACATCAAGAGATT	339	eSerSerIleAlaProValIleValProThrValSerGlyPheIleuLeu	52576	GTCAG 52572	357
52576	GTCAG 52572	357	52576	GTCAG 52572	357	52576	GTCAG 52572	357

[illegible]

LOCUS 140028 2716 bp DNA PAT 13-MAY-1997
 DEFINITION Sequence 1 from patent US 5618677.
 ACCESSION 140028
 VERSION 140028.1 GI:2083033
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2716)
 AUTHORS Ni, B. and Paul, S.M.
 TITLE Human brain sodium dependent inorganic phosphate cotransporter assay
 JOURNAL Patent: US 5618677-A 1 08-APR-1997;
 FEATURES
 source Location/Qualifiers
 1. 2716
 BASE COUNT 504 a 854 c 738 g 620 t
 ORIGIN

alignment_scores:

quality: 479.00 length: 397
 Ratio: 2.056 Gaps: 7
 Percent Similarity: 58.690 Percent Identity: 28.212

alignment_block:

US-09-391-958-1 x 140028 ..

Align seg 1/1 to: 140028 from: 1 to: 2716

12 ValProSerLeuGysSerAlaArgTyrGlyLeuValLeuHisph 28
 :::::::::: :::::::::::::: :::::
 941 ATCCCTCAGCTGCGGCTGCTATGCGTGCATC..... 979
 28 eCysAsnPhrThrIleAlaGlnAsnValIleMetAsnIleThrMet 45
 ::::: ::::: ::::: ::::: ::::: :::::
 980TTGCTGAGATCCTGCAGGGGTTGGTAGAGGGGTCACATACC 1022
 45 AlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSer 61
 :::::
 1023 CCGCGCTGC..... 1030
 62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLys 78
 1030 1030
 78 rLeuProAlaLysSerSerIleLeuGlyGlnPheAlaIleTrpGlu 95
 :::::::::::::: :::::::::::::: :::::
 1031CATGGATCTGAGCA 1046
 95 rGTrpGlyProProGlnIuArgSerArgLeuGysSerIleAlaLeuSer 111
 :::::::::::::: :::::::::::::: :::::
 1047 AATGGGCCCACTTAGAAGAGTGGCTGGCGACGACAGCCTTTTGT 1096
 112 GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyPheIle 128
 ::::: ::::: ::::: ::::: ::::: :::::
 1097 GGTTCCTATGCGGGGGGCTGGTGGATGCCCTGGCGGGGCTCTGT 1146
 128 rGluThrLeuGlyTrpProPheValPheTyrIlePheGlyGlyAla 145
 ::::: ::::: ::::: ::::: ::::: :::::
 1147 GCAGTACTGAGATGAGCTGTTTCTACGCTACGACGACCTCGGGA 1196
 145 yValAlcysLeuLeuTrpPheValValIleTyrAspAspProValSer 161
 :::::::::::::: :::::::::::::: :::::
 1197 TCTTCTGACTGTCTGCTGCTGCTCTCTACGAGCCGCCCGCTG 1246
 162 TyrProTrpIleSerThrSerGluLysGluTyrIleIleSerSerLeu 178
 :::::::::::::: :::::::::::::: :::::
 1247 CACCCAGCATCTCGAGAGAGAGGACAGTACATC.....GA 1284
 178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188
 :::::::::::::: :::::::::::::: :::::
 1285 GCACGCCATCGAGAGAGCGCAAACTCATGAAACCCCTCAGAGATTTA 1334

189ProIleLysAlaMetLeuArgSerLeuProIleTrpSerIleCys 203
 ::::: ::::: ::::: ::::: ::::: :::::
 1335 GCACCTCCCTGGCGGCTTTCACGTTCATGACAGTCAATGACATCATC 1384
 204 LeuGlyCysPheSerHisGlnTrpLeuValSerThrMetValValTyr 220
 ::::: ::::: ::::: ::::: ::::: :::::
 1385 GTGGCCAACTTCTCCAGCTGAGCTTCTACCTGCTGCTCATCTCCA 1434
 220 eProThrTyrIleSerSerValTyrHisValAsnIleArgAspAsnGly 237
 ::::: ::::: ::::: ::::: ::::: :::::
 1435 GCCCGACTACTTCGAGAAGGTTCGCGCTTCGAGATCAGCAAGTAGGCC 1484
 237 LeuLeuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
 ::::: ::::: ::::: ::::: ::::: :::::
 1485 TGGTGTCCGGCTGCCACCTGCTCATGACCATCATCGTCCCATCGGC 1534
 254 GlyTyrLeuAlaAspPheLeuThrLysLysPheArgLeuIleThr.. 269
 ::::: ::::: ::::: ::::: ::::: :::::
 1535 GCCCAGATCCGCGACTTCTGCGAGCGCGCATCATGTCACCAACCA 1584
 270 .ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 286
 ::::: ::::: ::::: ::::: ::::: :::::
 1585 CGTGGCGAAGTGTATGAACTCGCGAGGCTTCGGCATGAGAACCCAGCTGC 1634
 286 IeValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAlaLeu 302
 ::::: ::::: ::::: ::::: ::::: :::::
 1635 TGTGTGTGCTGGCTACTCGACTCCAGAGGCGTG...GCCATCTCTTTC 1681
 303 LeuThrLeuSerCysGlyLeuSerThrLeuGysGlnSerGlyLeuTyr 319
 ::::: ::::: ::::: ::::: ::::: :::::
 1682 CTGTGCTTACCGCTGGCTTCACCGGCTTCGCCATCTGTGGTTCAACGT 1731
 319 eAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAla 336
 ::::: ::::: ::::: ::::: ::::: :::::
 1732 GAACCACTGAGACATAGCCCGCGCTACCGCAGCATCTCATGAGGATCT 1781
 336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly 352
 ::::: ::::: ::::: ::::: ::::: :::::
 1782 CCACAGCGCTGGGACACACTGTGGCATGTGTGCCCATCATCGTGGGG 1831
 353 PheLeuLeuSerGlnAspProGluPheGlyTyrPArgAsnValPhePhe 369
 ::::: ::::: ::::: ::::: ::::: :::::
 1832 GCCATGACTAAGACACAAAGCTCGGAGGAGTGGCATGAGTGTCTTAT 1881
 369 uLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrIlePheGly 386
 ::::: ::::: ::::: ::::: ::::: :::::
 1882 TGCCTCCTGCTGCACTATGAGGTGTCACTTCTACGCGGCTTTCCT 1931
 386 LuAlaAspValGlnIuTrpAlaLysGluArgLysLeuThr 399
 ::::: ::::: ::::: ::::: ::::: :::::
 1932 CTGGAGAGAGACAGCCGTTGGCGAGGCTGAGAGATGAGC 1972
 seq_name: gb-pat:140029
 seq_documentation_block:
 LOCUS 140029 2716 bp DNA PAT 13-MAY-1997
 DEFINITION Sequence 3 from patent US 5618677.
 ACCESSION 140029
 VERSION 140029.1 GI:2083034
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2716)
 AUTHORS Ni, B. and Paul, S.M.
 TITLE Human brain sodium dependent inorganic phosphate cotransporter assay
 JOURNAL Patent: US 5618677-A 3 08-APR-1997;
 FEATURES
 source Location/Qualifiers
 1. 2716
 BASE COUNT 504 a 854 c 738 g 620 t

ORIGIN

alignment_scores:

Quality: 479.00 Length: 397
Ratio: 2.056 Gaps: 7
Percent Similarity: 58.690 Percent Identity: 28.212

alignment_block:

US-09-391-958-1 x I40029 ..

Align seg 1/1 to: I40029 from: 1 to: 2716

```
12 ValProSerLeuCySerAlaArgTyrGlyIleAlaLeuValLeuHisph 28
   ::::::::::::::::::::::::::::
941 ATCCCTCAGCTGCCCGCTCCTAGCGCTGCTCATC..... 979

28 eCysAsnPheThrIleAlaGlnAsnValIleMetAsnIleThrMetV 45
   |||::: ||| ::::::::::::::: ::::::::::
980 .....TTCGTGAGGATCCTCGACAGGGGTGTGAGGGGCTCACATACC 1022

45 aAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
   |||
1023 CCGCCTGC..... 1030

62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLysSe 78
1030 ..... 1030

78 rLeuProAlaLysSerSerIleLeuGlyGlyLnpheAlaIleTrrGlnA 95
1031 .....CATGGGATCTGGAGCA 1046

95 rGrrPGlyProProGlnGlnArgSerArgLeuCySerIleAlaLeuSer 111
   :::::::::::::: ||| ||| ||| ||| ||| ||| ||| ||| |||
1047 AATGGGCCCCACCTTGAACGAGATCGCTGCGGACGACAGACCTTTGT 1096

112 GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyIlePheIleSe 128
   ||| ||| ::::::::::: ::::::::::
1097 GGTTCCTAATGCTGGGCGGTGGTCCGATGCCCTCGCCGGGTCCTTGT 1146

128 rGluThrLeuGlyTrrProPheValPheTrrIlePheGlyGlyValGlyC 145
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1147 GCAGTACTCAGAGATGGAGCTCTGTTTCTACGTCTACGGCAGCTTCGGGA 1196

145 ysValCysCysLeuLeuTrrPheValValIleTyrAspAspProValSer 161
   ||| ::::::::::: ||| ||| ||| ||| ||| ||| ||| |||
1197 TCTCTGTGATACCTGTCTGCTGCTGCTCTCTACGAGTCCCGCGCTG 1246

162 TyrProTrrIleSerThrSerGlnLysGluTrrIleIleSerSerLeuLys 178
   ::||| ||| ||| ||| ::::::::::: ||| ||| ||| |||
1247 CACCCCAACATCTCGAGAGAGGAGCGCAATGATC.....GA 1284

178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188
   ::::: ::::::::::||| :::: ||| ||| ||| ||| ||| |||
1285 GGAGGCCATCGAGAGAGCGGAAACATCATACCCCTCAGAAATTGA 1334

189 .....ProIleLysAlaMetLeuArgSerLeuProIleTrrSerIleCys 203
   ||| ::::: :::::::::: ||| ||| ||| ||| ||| ||| |||
1335 GCACCTCCCTGGCGGGCTTTCACGTCTATGCCAGTATGCCATCATVC 1384

204 LeuGlyCysPheSerHisGlnTrrPheValSerThrMetValValTrrI 220
   :::::: ||| ::::::::::|||
1385 GTGGCCAACTTCTGCCGAGCTGAGCTTACCTGCTCATCTCCCA 1434

220 eProThrTrrIleSerSerValTrrHisValAsnIleArgAspAsnGlyL 237
   ||| ||| ::::::::::||| :::: ||| ||| ||| ||| ||| |||
1435 GCCCGACTACTTCCAGAAAGTGTTCGGCTTCGAGATCGCAAGTAGGCC 1484

237 euleuSerAlaLeuProPheIleValAlaTrrValIleGlyMetValGly 253
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1485 TGGTGTCCGCGCTGCCACCTGTGATGACATCATCGTCCCATCGGC 1534
```

```
254 GlyTrrLeuAlaAspPheLeuLeuThrLysLysPheArgLeuIleThr.. 269
   ||| ::::::::::||| ||| ::::::::::: |||
1535 GGCCAGATCGCGGACTTCTTCGGAGCCGCCGATCATGTCCACCACCAA 1584

270 .ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 286
   ||| ||| ||| ::::: ||| ::::::::::: |||
1585 CGTGGCAAGTTGATGAACTCGGAGGCTTGGGCTGGAGAACCCAGCGTGC 1634

286 lValSerLeuProTrrIleuAsnSerGlyTrrIleThrAlaThrAlaLeu 302
   ::::: :::: ||| ::||| ::::: ||| ::::: |||
1635 TGTGTGTGTGCTGCTACTCGCATCTCCAAGGGCGTGC...GCCATCTCTTC 1681

303 LeuThrLeuSerCysGlyLeuSerThrLeuGlyGlnSerGlyIleTrrI 319
   ||| ||| ||| ::::: ||| ::::: ||| ||| ||| ||| |||
1682 CTGGCTTACCGCTGGGCTTCAGCGGCTTCCCATCTCTGGGATCAACGT 1731

319 eAsnValLeuAspIleAlaProArgTrrSerSerPheLeuMetGlyAlas 336
   ::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1732 GAACCACTTGACATATAGCCCGCGCTACGCCAGCATCTCTCATGGGCATCT 1781

336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly 352
   ||| ||| ::::::::::: ::::: ||| ::::: |||
1782 CCAACGGCGGTGGCACACTGTTCGGGCATGTGTGCCCATCATCTCGTGGG 1831

353 PheLeuLeuSerGlnAspProGlnPheGlyTrrPargAsnValPhePheLe 369
   ::::: ::::::::::: ||| ||| ||| ||| ||| ||| |||
1832 GCCATGACTAAGCACACAGACTCGGGAGAGTGGCATGTGCTCTATAT 1881

369 uLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTrrLeuIlePheGlyG 386
   :: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1882 TGCCCTCCCTGGTGCATGTGAGAGTGTATCTTCTACGGGCTCTTGCTT 1931

386 lValAspValGlnGluTrrPalaLysGluArgLysLeuThr 399
   :::::: ||| ||| ||| ||| ::::::::::: |||
1932 CTGGAGAGAACACAGCCGTGGGACAGAGCTGTGAGAGATGAGC 1972
```


OM of: US-09-391-958-1 to: N_Geneseq_36:* out_format : pbs
Date: May 25, 2000 1:59 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=firmet p2n.model -DEFV=xlp
-O=/c9q2_1/USFTO.spool/US09391958/rnatc_24052000_213330_7207/app-query.fasta.1
-DB=N_Geneseq_36 -QFPT=fastcap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blcsunm2
-TRANS=human40.ctd -LIST=45 -DOCCALIGN=200 -THR_SCORE=pct
-ALIGN=15 -MODE=LOCAL -OUTEXT=pbs -NORM=ext -MINLEN=0
-MAXLEN=1000000 -USER=US09391958 -NCPU=6 -ICPU=3 -NO_XLPPXV -WAIT
-THREADS=1

Search information block:
Query: US-09-391-958-1
Query length: 401
Database: N_Geneseq_36:*
Database sequences: 311585
Database length: 125096042
Search time (sec): 62.680000

score_list:

Sequence	Strnd	Orig	ZScore	Escore	Len	Documentation
N_Geneseq_36:V43711	+	2076.00	3760.06	7.8e-202	1643	Human sodium-dependent phosphatase
N_Geneseq_36:V57910	+	2061.00	3731.66	3.0e-200	1795	Human haemochromatosis gene NT
N_Geneseq_36:V57909	+	932.00	1667.68	2.7e-85	2281	Human haemochromatosis gene NT
N_Geneseq_36:V42064	+	479.00	838.63	4.1e-39	2916	Human brain Na+ dependent inot
N_Geneseq_36:V33503	+	479.00	838.63	4.1e-39	2716	Human sodium-lithium counteri
N_Geneseq_36:V43712	+	433.00	781.08	6.6e-36	2732	Sodium-dependent phosphate cotr
N_Geneseq_36:V57926	+	318.00	493.46	6.9e-20	235033	Hereditary haemochromatosis
N_Geneseq_36:V57903	+	318.00	493.46	6.9e-20	237326	Hereditary haemochromatosis
N_Geneseq_36:V52652	+	169.50	271.13	5.3e-09	1840	DNA encoding protein involved
N_Geneseq_36:V52653	+	169.50	271.13	5.3e-07	3217	DNA encoding protein involved
N_Geneseq_36:V74423	+	150.50	239.49	9.7e-06	2576	Staphylococcus aureus contig S
N_Geneseq_36:V74423	+	138.50	194.87	0.0030	18613	Staphylococcus aureus contig S
N_Geneseq_36:V52165	+	131.00	188.04	0.0071	10240	Staphylococcus aureus contig
N_Geneseq_36:V74708	+	124.00	212.08	0.0003	415	Staphylococcus pneumoniae gen
N_Geneseq_36:X07357	+	122.50	195.49	0.0027	5385	Staphylococcus aureus contig SB
N_Geneseq_36:X13329	+	120.50	176.25	0.0323	2288	Enterococcus faecalis genome C
N_Geneseq_36:X13024	+	118.50	182.43	0.0146	32768	Enterococcus faecalis genome C
N_Geneseq_36:X13060	+	115.50	146.38	1.49	3225	Enterococcus faecalis genome
N_Geneseq_36:X13233	+	113.50	169.36	0.0783	3181	Enterococcus faecalis genome G
N_Geneseq_36:V47611	+	112.50	167.69	0.0969	9271	Staphylococcus aureus contig S
N_Geneseq_36:V23080	+	112.50	155.41	0.4686	10974	Fragment HcJ1741 of a new Hepa
N_Geneseq_36:N60055	+	112.50	153.47	0.6007	2584	Sequence of VB131 clone circu
N_Geneseq_36:V51330	+	112.00	169.17	0.0802	7296	Human Na+/H+ exchanger isoform
N_Geneseq_36:X12996	+	112.00	157.25	0.3701	9333	Enterococcus faecalis genome C
N_Geneseq_36:V23084	+	111.50	153.51	0.5980	9256	Fragment F1029 of a new Hepat
N_Geneseq_36:V23077	+	110.50	151.78	0.7467	9391	Fragment HcJ1741 of a new Hepa
N_Geneseq_36:V23075	+	109.50	151.61	0.7628	9034	Fragment HcJ1741 of a new Hepa
N_Geneseq_36:T04247	+	109.50	150.23	0.9105	9122	Hepatitis GB virus (HGBV) cont
N_Geneseq_36:T42920	+	109.50	150.12	0.9236	9122	GB-C viral genomic sequence. H
N_Geneseq_36:T76929	+	109.50	149.78	0.9638	9390	GT230 genomic RNA sequence for
N_Geneseq_36:T76931	+	109.50	149.78	0.9639	9390	GT230 gene cDNA sequence for c
N_Geneseq_36:V52179	+	109.00	151.34	0.7900	2885	Staphylococcus pneumoniae genom
N_Geneseq_36:T35417	+	107.50	159.46	1.48	9126	Human amine transporter cDNA.
N_Geneseq_36:T45828	+	107.50	146.46	1.48	9126	Hepatitis GB virus isolate C. 1
N_Geneseq_36:T96447	+	107.50	146.46	1.48	9126	Fragment of HGBV NS3 genomic R
N_Geneseq_36:V39074	+	107.50	146.46	1.48	9126	Fragment of HGBV NS3 genomic R
N_Geneseq_36:V23076	+	107.50	146.13	1.54	9391	Hepatitis GB virus E2 gene reg
N_Geneseq_36:T76930	+	107.50	146.13	1.54	9391	Fragment HcJ1757 of a new Hepa
N_Geneseq_36:T76930	+	107.50	146.13	1.54	9395	GT110 gene cDNA sequence for c
N_Geneseq_36:T58840_0	+	107.50	117.87	57.77	110000	Mycoplasma genitalium genom
N_Geneseq_36:V23082	+	107.00	146.19	1.53	8630	Fragment HcJ1775 of a new Hepa
N_Geneseq_36:V12391	+	106.50	161.50	0.2146	2102	Mouse osteoclast transporter p

N_Geneseq_36:T15656 + 106.50 144.31 1.95 9391 | HGV-PNF 2161 polyprotein co
N_Geneseq_36:T94168 + 106.50 144.31 1.95 9391 | Hepatitis G virus PNF-2161
N_Geneseq_36:V23083 + 106.50 144.31 1.95 9391 | Fragment K3732(21) of a new
N_Geneseq_36:T08812 + 106.50 144.31 1.95 9392 | Hepatitis virus Open readin
seq_name: N_Geneseq_36:V43711

seq_documentation_block:
ID V43711 standard; DNA: 1643 BP.
AC V43711:
AT 30-NOV-1998 (first entry)
DE Human sodium-dependent phosphate cotransporter coding sequence.
KW Human sodium-dependent phosphate cotransporter; human; NAFTR; cancer; myopathy;
KW cell signalling disorder; phosphate regulation disorder; therapy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 235..1440
FT /tag= a

MO9837198-A1.
PD 27-AUG-1998.
PF 24-FEB-1998; U03745.
PR 24-FEB-1997; US-805118.
PA (INCY-) INCYTE PHARM INC.
PI Bandman O, Lai P;
DR WPI: 98-467561/40.
DR P-PSDB: W69971.
PT New isolated human sodium-dependent phosphate co:transporter - used
PT to develop products for treating e.g. cancers, osteoporosis,
PT Alzheimer's disease, diabetes, encephalopathy, myopathy,
PT hypocalciuria or hypoglycaemia.
PS Claim 5: Fig 1: 66pp: English.

This sequence encodes the human sodium-dependent phosphate cotransporter
(NAFTR) of the invention. NAFTR and agonists of it can be used to treat
or prevent disorders associated with decreased phosphate levels,
e.g. cancers of the kidney, disorders of decreased phosphate levels
including tumoral calcinosis, osteomalacia, osteoporosis, familial
hypophosphataemia, rickets, cystenemia, nephrocalcinosis,
glomerulonephritis, renal calculus, Alzheimer's disease, diabetes
mellitus, hereditary amyloidosis, myopathies including progressive
external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic epilepsy,
encephalopathy, cardiomyopathy, hypokalaemia, Goodpastures syndrome, and
disorders of cell signalling through cAMP, ATP, NADPH and
glucose-6-phosphate. Antagonists or inhibitors of NAFTR may be
administered to a subject to treat or prevent disorders associated with
increased phosphate levels, e.g. hypocalciuria, hypocalcaemia, and
abnormal phosphate regulation in neurons, gastrointestinal tract and
liver. The products can also be used for detection, diagnosis and drug
screening.

Sequence 1643 BP; 435 A; 376 C; 359 G; 473 T;
alignment_scores:
Quality: 2076.00 Length: 401
Ratio: 5.177 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-391-958-1 x V43711 ..
Align seg 1/1 to: V43711 from: 1 to: 1643

1 MetGlnValSpGluThrLeuAlaIleProAlaGlyValProSerLeuGlyCys 17
ATTCAGATGATGATGACACATGATCCCGAGAAAGTTCCAGATTATGTTCC 284
235 ATTCAGATGATGATGACACATGATCCCGAGAAAGTTCCAGATTATGTTCC 284
17 TAlaArgTyrGlyTlleAlaLeuValleuHisPheCysAsnPhenThrTrp 34
TCTTCGCTTATGCAATACCCCTCGCTTACATTTCTGCAATTTTCAACAGCA 334
285 TCTTCGCTTATGCAATACCCCTCGCTTACATTTCTGCAATTTTCAACAGCA 334
34 TleAlaGlnAsnValIleMetAsnIleThrMetCysValAlaMetValAsnSer 50
TATGCAAAATGATCATCATGACATGACATGACATGACATGACATGACATGAC 384
335 TATGCAAAATGATCATCATGACATGACATGACATGACATGACATGACATGAC 384

51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
 385 ACAAGCCCTCAATCCAGACTCAATGATTCCTGAGGTGGCTGTTGA 434
 67 PsePhGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSers 84
 435 CTCATTGGGGCTTAAGTAAAGCCCAAGAGCTTCCTGCAAGTCCT 484
 84 erileuGlyGlyGlnPheAlaIleTyrGluAspGlyProGln 100
 485 CAATACCTGGGGGCTCACTTTCGCAATTTGGGAAAGGTGGGCTCCACA 534
 101 GluArgSerArgLeuGlySerLysAlaLeuSerGlyMetLeuGlyGly 117
 535 GAACGAGAGAGACTCTGACATGCTTTATCAGAAATGTTACTGGGATG 584
 117 SPheThrAlaIleLeuIleGlyPheIleSerGlyTrpThrLeuGlyTrp 134
 585 CTTTACTGCCATCCTCATAGGTGGCTCATAGTAAAGCCCTGGGTGGC 634
 134 robPheValPheTyrIlePheGlyGlyValGlyCysValCysCysLeu 150
 635 CCTTGGCTCTCATATCTTGGAGGTGGTGGCTGCTGCTGCTTCTC 684
 151 TrpPheValValIleTyrAspAspProValSerTyrProTyrIleSer 167
 685 TGGTTTGGTGTGATTATGATGAGCCCTTCTCATCAGATGATTAAGC 724
 167 PserGlyLysGlyTyrIleIleSerSerLeuGlyGlnGlnValGlySers 184
 735 CTCAGAAAAAGATACATCATATCTCTTGAACAAACAGGTGGGTCTT 784
 184 erlyGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTyr 200
 785 CTAAAGAGCTCTCCATCAAAAGATGCTCAGATCTACCATTTGG 834
 201 SerIleCysLeuGlyCysPheSerHisGlnTyrPheValSerThrMet 217
 835 TCCATATGTTAGGCTGTTTACCCATCAATGATTGTTAGCAAAATGCT 884
 217 ValTyrIleProThrTyrIleSerSerValTyrHisValAsnIleArg 234
 885 TGTATACATACCACTTACATCAGCTCTGTACCATGTTAACATCAGAG 934
 234 SPasnGlyLeuLeuSerAlaLeuProPheIleValAlaTyrValIleGly 250
 935 ACAATGAGCTTCTATCTGCCCTTCTTATTTGTTGCCGTGATAGAGC 984
 251 MetValGlyGlyTyrIleValAspPheLeuLeuThrLysLysPheArg 267
 985 ATGGTGGAGGCTATCTGCAATTTCTCTTACCAAAAAGTTTAGACT 1034
 267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSer 284
 1035 CATCACTGTGAGGAAATGCAACATTTTAGAAGTCCCTCTTTCAG 1084
 284 IaleuIleValSerLeuProTyrIleAsnSerGlyTyrIleThrAlaThr 300
 1085 CACTATATGTCTCTGCTTACCTCAATCCGGCTATATACACCAACT 1134
 301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGly 317
 1135 GCCTTGTGAGGCTCTCTTGGGATTAAGACATTTGTGACATCGGAGT 1184
 317 eTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMet 334
 1185 TTAATATCAATGCTTATGATATGCTCAAGGATATCCAGTTTCTCATGG 1234
 334 LAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
 1235 GAGCATCAAGAGATTTTCAGACATAGCACTGTCATTGTCACCATCTTC 1284
 351 SerGlyPheLeuLeuSerGlnAspProGlnPheGlyTyrPargAsnValPh 367

seq_name: N_Geneseq_36:V57910
 seq_documentation_block:
 ID V57910 standard; cDNA; 1795 BP.
 AC V57910.
 DT 21-DEC-1998 (first entry)
 DE Human haemochromatosis gene NTP4.
 KW Bovine butyrophilin; BR; human hereditary haemochromatosis; HFE;
 KW diagnosis; iron metabolism; NPT3; NPT4; Roret; BTF1; BTF2; BTF3;
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
 KW type 1 sodium transport gene; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT 377..1582
 FT CDS /tag= a
 FT /product= "NTP4"
 PN W09814466-A1.
 PD 09-APR-1998.
 PR 30-SEP-1997; U17658.
 PR 07-MAY-1997; US-852495.
 PR 01-OCT-1996; US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kromal GS, Laufer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 DR P-PSDB: W78920.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Claim 58: Fig 7: 209pp. English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the haplotype or genotype indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophilin (BTF), and can be used in the production of agonists
 CC and antagonists of BTF function. Also described are: (1) a Roret gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence encodes
 CC NTP4.
 SQ Sequence 1795 BP; 473 A; 436 C; 402 G; 484 T;
 alignment_scores:
 Quality: 2061.00 Length: 401
 Ratio: 5.165 Gaps: 0
 Percent Similarity: 99.501 Percent Identity: 99.252
 alignment_block:
 US-09-391-958-1 x V57910 ..

Align seg 1/1 to: V57910 from: 1 to: 1795

```

1 MetGlnValAspGluThrLeuIleProArgLysValProSerLeuGlySse 17
  |||||||
377 ATGCAAGATGATGACACTGATCCCGAGGAAAGTCCAGATTATGCTTC 426
17 AlaAArgTyrGlyIleAlaLeuValLeuHisPheCysAsnPheThrThr 34
  |||||||
427 TGTGCTGCTATGGAATAGCCCTGCTTACATTCTGCAATTTACAAACA 476
34 LeuAlaGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
  |||||||
477 TACACAAAATGTCATCATGACATCACCATGTCAGCATGTCACACAGC 526
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
  |||||||
527 ACAAGCCCTCAATCCAGCTCAATGATTCCTCTGAGTGCCTGCTGTGA 576
67 PserPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSerS 84
  |||||||
577 CTCATTGGTGGCTAAGTAAGCCCCAAAGAGTCTTCTGCAAAAGCTC 626
84 eRIleLeuGlyGlyGlnPheAlaIleTrpGluArgTrpGlyProGln 100
  |||||||
627 CAATACTTGGGGGTCAAGTTTGCATTTGGGAAAGTGGGGCCCTCCACA 676
101 GluArgSerArgLeuGlySerIleAlaLeuSerGlyMetLeuLeuGlyCy 117
  |||||||
677 GAACGAGAGAGACTGTCAGCATTCCTTATCAGGAATGTTACTGGGATG 726
117 sPheThrAlaIleLeuIleGlyLysPheIleSerGluThrLeuGlyTrp 134
  |||||||
727 CTTTACTGCCATCTCATAGTGGTTCATTAGTGAACCTTGGGGTGGC 776
134 roPheValPheTyrIlePheGlyIleValGlyCysValCysCysLeuLeu 150
  |||||||
777 CCTTGTCTCTCATATCTTTGGAGGTGGTGGCTGCTGCTGCTCTCTC 826
151 TrpPheValIleTyrAspAspProValSerTyrProTrpIleSerThr 167
  |||||||
827 TGGTTGTGTGATTTATGATGATGACCCCTTCTCATCCATGAGATAGAC 876
167 rSerGluLysGluTyrIleIleSerSerLeuLysGlnGlnValGlySerS 184
  |||||||
877 CTGAGAAAAGATACATCATCTCTCTGAAACAACAGCTGGGGCTT 926
184 eRlySGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrp 200
  |||||||
927 CTAGAGCAGGCTCTCCCATCAAGCATGCTCAGATCTCAACCATTTGG 976
201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMetVa 217
  |||||||
977 TCCATATGTTTAAAGCTGTTTACCCATCAATGATGTTAGTGTAGCACA 1026
217 ValTyrIleProThrTyrIleSerSerValTyrHisValAsnIleArg 234
  |||||||
1027 TGTATACATACCAACTTACTACGCTCTGTGACCAATGTTAAACATCAG 1076
234 sPasnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
  |||||||
1077 ACAATGGACTTCATCTGCCCCCTCTTTATTTTGTGCTGGGTCATAGGC 1126
251 MetValGlyGlyTyrLeuAlaAspPheLeuLeuThrLysLysPheArgLe 267
  |||||||
1127 ATGGTGGGAGGCTATCTGCGCAGATTTCTTCTTAACCAAAAAGTTAGACT 1176
267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSera 284
  |||||||
1177 CATCACTGTGAGGAAATGCCACAAATTTTAAAGAAAGCTCCCTCTTCAG 1226
284 IalaLeuIleValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThr 300
  |||||||

```

```

1227 CACTCATTTGTGCTGCTGCTTACCTCAATTCGGGCTATATCACAGCACT 1276
301 AlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuGlnSerGlyI 317
  |||||||
1277 GCGTTGCTGACGCTCTCTTCCGATTAAGCACAATGTGTACAGTCAGGAT 1326
317 eTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMet 334
  |||||||
1327 TTATATCAATGCTCTTAGATATGTCTCCAAAGTATTCAGATTTCCTCAG 1376
334 lValaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
  |||||||
1377 GAGCATCAAGAGATTTTTCAGCATAGCACCTGTCATTTGACCACTGTC 1426
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTyrTrpArgsnVal 367
  |||||||
1427 AGCGGATTTCTTCTTAGTACAGACCCGAGATTTGGGGAGGAAATGCTT 1476
367 ePheLeuLeuPheAlaValAsnLeuGlyLeuLeuPheTyrLeuIleP 384
  |||||||
1477 CTCCTTGTCTGTGCGCTTAACCTGTAGGACTACTCTTCTTACCTCATAT 1526
384 heGlyGlnAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrArg 400
  |||||||
1527 TTGGAGAGAGAGATGTCCAGAAATGGGCTAAAGAGAAACATCACTCCT 1576
401 Leu 401
  |||
1577 TTTA 1579

seq_name: N_Geneseq_36:V57909
seq_documentation_block:
ID V57909 standard; cDNA; 2281 BP.
AC V57909;
DR 21-DEC-1998 (first entry)
DE Human haemochromatosis gene NRP3.
KW Bovine butyrophilin; Bt; human hereditary haemochromatosis; HFE;
  diagnosis; iron metabolism; NRP3; NRP4; Rokit; BTF2; BTF3;
  BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
  type 1 sodium transport gene; ss.
OS Homo sapiens.
FH Key location/Qualifiers
FT CDS 419..1729
FT FT /*tag= a
FT FT /product= "NRP3"

WO9814466-A1.
09-APR-1998.
30-SEP-1997: U17658.
07-MAY-1997: US-852495.
01-OCT-1996: US-724394.
(Patent) PROGENITOR INC.
PI Feder JN, Krommal GS, Lauer PM, Ruddy DA, Thomas WJ,
  Tsuchihashi Z, Wolff RK;
PI WPI: 98-240014/21.
DR P-PSDB: W78919.
DR Hereditary haemochromatosis gene products - used to develop products
  for the diagnosis and treatment of hereditary disorders in iron
  metabolism
PT Claim 52: Fig 7: 209pp; English.
PS The present invention describes hereditary haemochromatosis gene
  products from the human haemochromatosis gene. Also described is a
  method to determine the presence or absence of the common hereditary
  haemochromatosis (HFE) gene mutation in an individual comprising:
  (a) providing DNA or RNA from the individual; and (b) assessing the
  DNA or RNA for the presence or absence of a haplotype or genotype where
  the presence or absence of the haplotype genotype indicates the likely
  presence of the HFE gene mutation in the genome of the individual. The
  HFE gene sequences from the present invention can be used to develop
  products for use in the diagnosis and treatment of HFE. The present
  invention also describes BTF genes, which are homologues of the milk
  protein butyrophilin (Bt), and can be used in the production of agonists
  and antagonists of Bt function. Also described are: (1) a Rokit gene

```

CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphatemia. The present sequence encodes
 CC NTP3.

Sequence 2281 BP: 624 A; 506 C; 490 G; 661 T;

alignment_scores:

Quality: 932.00 Length: 471
 Ratio: 3.305 Gaps: 4
 Percent Similarity: 59.873 Percent Identity: 41.826

alignment_block:

US-09-391-958-1 x V57909 ..

Align seg 1/1 to: V57909 from: 1 to: 2281

```

10 ArgLysValProSerLeuSerAlaArgTyrGlyIleAlaLeuValle 26
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 AGGAAAGGTCACAGATTCTGTCATTACGCTATGAGCTGGCTTTATCAT 489
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
26 uHisPheCysAsnPheThrThrIleAlaGlnAsnValIleMetAsnIleT 43
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 GCACCTTCTCAACTTCACCTTCACATGATTAAGCAGCGTGTAGCTGACGAT 539
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
43 hMetValAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAsp 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 CGATCATCCCATGTCGTGAACACACACACACAGCAGCAGTATCTATATGCC 589
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
60 SerSerGluValLeuProValAspSerPheGlyGlyLeuSerLysAlaPr 76
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 TCCACTGAGGGGCGCTGTGTGACAGATGCTTCAATTAACCTCCAGATATCCAT 639
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
76 oLysSerLeuProAlaLysSerSerLe.. 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 CAAGGAATTTGATPACAAAGGCTCTGTGTATCATGAGGCCAGAAATCTC 689
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
85 .. 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
690 AGGGATCATCTTATGCTTCATCACTAATGAGATATACTACTCTGATTC 739
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
86 .. 86
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
740 CCAATGGATATTAGACAGCGATATTGGAGCAAAAAAATGCTGGTGC 789
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
87 .. 87
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
790 TGGTTGCTGATCTTCCCTTCTCACCCTCTTTACACACTGGCTGCTG 839
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
87 .. 87
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
840 ACTTGGAGTATTTGGTTCATCATGTTCCGACATCCAGGGCATGGCC 889
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
88 .. 88
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
890 CAGGAAATGGCATGACAGGTCACTTACTATTGGGCAAAAGTGGGCTGC 939
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
98 oProGlnIleArgSerArgLeuSerSerIleAlaLeuSerGlyMetLeuL 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
940 TCCACTTGAAGCAAGACAGCTCACACCATTCAGAGATCAGGTCAGCAT 989
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
115 euGlyCysPheThrAlaIleLeuIleGlyGlyPheIleSerGluThrLeu 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
990 TTGGATCTCTATCATCTCTGTGTGGGGGACATTAATCTCAGAGCCCTTG 1039
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
132 GlyTrpPropheValPheThrIlePheGlyGlyValGlyCysValCysC 148
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1040 AGCTGGCCTTTATCTTCTATCATCTTTGGTGAACACTGGCTGTGCTGCTG 1089
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
148 sleuLeuThrPheValValIleTyrAspAspProValSerTyrProTPr 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1090 TCTCTATGTTGTCACAGTATGATTAATGATGACCCATCATCACCCGTGCA 1139

```

```

165 leSerThrSerGlyLysGlyTyrIleIleSerSerLeuSerGlnGlnVal 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1140 TAAGTGTAGGGAAGAAAGAGACATCTCTCCCTACCTGCTCAACAGGCC 1189
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
182 GlySerSerLysGlnProLeuProIleLysAlaMetLeuArgSerLeuPr 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1190 AGTTCTCTGAGACGAGCTGCCCATTAAGGCGATGTCACATGCTTACC 1239
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
198 oLleTrpSerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerT 215
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1240 ACTTGGGCGCATTTCTCTGCTGCTTTTTCAGCCATTTCTGTGTATGACCA 1289
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 hMetValValIleThrIleProThrTyrIleSerSerValTyrHisValAsn 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1290 TCATCTTAACATACCTACCAACAGTATATCATGATCTGCTCATTTAAC 1339
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
232 IleArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpVa 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1340 ATCAGAGATAGTGAAGTCTGTCTCCCTGCTTTTATTTGCTGCTGCAAG 1389
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
248 IleGlyMetValGlyGlyTyrLeuAlaAspPheLeuThrLysLys. 264
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1390 CTGTACAATTTTAGAGAGTCACACTGCGACATTTCTTTTTCACAGGAATC 1439
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 .. PheArgLeuIleThrValAlaArgLysIleAlaThrIleLeuGlySerLeu 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1440 TTCTCAGATGTATCAGTGTGCGAAAGCTTTTTCATCTCTTGA..... 1482
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
281 ProSerSerAlaLeuIleValSerLeuProTyrIleAsnSerGlyTyrIle 297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 .. 1482
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
297 eThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuCysG 314
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 .. 1482
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
314 InsSerGlyIleTyrIleAsnValLeuAspIleAlaProArgTyrSerSer 330
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1483 .. 1483
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
331 PheLeuMetGlyAlaSerArgLysPheSerSerIleAlaProValIleVa 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1492 TTCTCATGGGAATCTCAAGGGATTTGGGCTCATTCGAGAAATCATCTC 1541
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
347 lProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGlyTyrPA 364
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1542 TTCACACTGCGCACTGATTCCTCATCAGTACAGATTTTGAAGTGTGTTGA 1591
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
364 rGAsnValPhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPhe 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1592 GGAATGCTTTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTT 1641
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
381 TyrLeuIlePheGlyGlyAlaAspValGlnIleTyrPalaLysGlyArgGly 397
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1642 TACCTTCACGTTTGGACAAGAGAACTTCAAGACATGGGCCAAAGAGAGAC 1691
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
397 sleuThrArgLeu 401
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1692 CTTTACCGGCTC 1704
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
seq_name: N_Geneseq_36:T42064
seq_documentation_block:
ID T42064 standard; cDNA; 2716 BP.
AC T42064;
DT 30-JAN-1997 (first entry)
DE Human brain Na+ dependent inorganic phosphate cotransporter cDNA.
KW Brain sodium-dependent inorganic phosphate cotransporter; hbnpt;
KW ss.
OS Homo sapiens.
FH Key
CDS Location/Qualifiers
FT 461..2143

```


PD 03-SEP-1998.
PR 11-FEB-1998; U02875.
PR 27-FEB-1997; US-039462.
PA (UYEM-) UNIV EMORY.
PI Gunn RB, Timmer RT;
DR WPI; 98-520759/44.
P-PSDB: W70500.
PT New isolated lithium-sodium counter-transporter DNA - used to
PT develop products for evaluating lithium-sodium transport in
PT erythrocytes, particularly for lithium therapy in manic depression.
PS Claim 7; Page 48-50; 64pp; English.
CC This DNA molecule encodes the human brain-specific lithium-sodium
CC countertransporter (LSTC) BNPI (see W70500). LSTCs provide the
CC physiological mechanism for the extrusion of lithium from cells,
CC i.e. they regulate the cell concentration of lithium. LSTC activity
CC determines the therapeutic effect of lithium. The invention
CC provides a simple molecular biological test for the ability of
CC cells to extrude lithium. LSTCs have significance for
CC determining the responsiveness of humans with mental disorders,
CC including manic depressives, to treatment with lithium salts.
CC Probes and primers for BNPI, P1T-1 (see V33501) P1T-2 (see
CC V33502) can be used in diagnostic tests useful for genetic
CC screenings to predict whether a patient will respond to lithium
CC treatment. The test is also a screen for susceptibility to, and
CC extent of, manic depressive illness, and is suitable for screening
CC newborns.

	Quality: 479.00	Length: 397
	Ratio: 2.056	Gaps: 7
	Percent Similarity: 58.690	Percent Identity: 28.212
Alignment block:		
US-09-391-958-1 x V33503 ..		
Align seg 1/1 to: V33503 from: 1 to: 2716		
12	ValProSerLeuGysSerAlaArgTYGlyLEAlaLeuValLeuHisPh	28
941	ATCCCTCATGCGCCGCGCCACATAAGCTGTGTATC.....	979
28	eGysAsnPhetIrrThIleAlaInsuValIleMetAsnIleThrMetV	45
980TTTCGAGAGATCCTCGAGGGGTGTGTAAGAGGGGTCAACAACC	1022
45	AlaMetValAsnSerIrrThSerProGlnSerIleuAsnAspSerSer	61
1023	CCGCGCTGC.....	1030
62	GlUValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLys	78
1030	1030
78	IleuProAlaLysSerSerIleLeuGlyGlyGlnPheAlaIleTrpGlu	95
1031CATGGATGTGGACGA	1046
95	rGTrpGlyProProGlnGluArgSerArgLeuGysSerIleAlaLeuSer	111
1047	AATGGGCCCCACCTTAGAACGAGAGTGCCTGGCAGACAGCCTTTTGT	1096
112	GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheIle	128
1097	GGTTCCTATGCTGGGGGGCGGTGTCGGAGAGCCCTTCGCGGGGTCTTGT	1146
128	rGluThrLeuGlyTrpProPheValPheTyrIlePheGlyGlyValGly	145
1147	GCAGACTCAGAGATGAGCTGTATTTCACGCTACGACGACGACTTGCGGA	1196
145	ysValGysCysLeuThrPheValValIleTyrAspAspProValSer	161

1197 TCTTTGGTACCGTTTCTGGCTGCTCTGCTCTACGAGTCCCGCGCTG 1246
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873

seq_name: N_Geneseq_36:V43712
seq_documentation_block:
ID V43712 standard; DNA; 272 BP.
AC V43712.
DT 30-NOV-1998 (first entry)
DE Sodium-dependent phosphate cotransporter coding sequence fragment.
KW Sodium-dependent phosphate cotransporter; human; NMPY; cancer; myopathia;
KW cell signalling disorder; phosphate regulation disorder; therapy; ss.
OS Homo sapiens.
PN M09837198-A1.

PD 27-AUG-1998.
 PF 24-FEB-1998: U03745.
 PR 24-FEB-1997: US-805118.
 PA (INCYTE) INCYTE PHARM INC.
 PI Bandman O, Lai P.
 DR WPI: 98-467561/40.
 PT New isolated human sodium-dependent phosphate co:transporter - used
 PT to develop products for treating e.g. cancers, osteoporosis,
 PT Alzheimer's disease, diabetes, encephalopathy, myopathy,
 PT hypocalciuria or hypoglycaemia.
 PS Example V; Page 47; 66pp; English.
 CC This sequence represents a fragment of the DNA encoding the human
 CC sodium-dependent phosphate cotransporter (NAPTR) of the invention. NAPTR
 CC and agonists of it can be used to treat or prevent disorders associated
 CC with decreased phosphate levels, e.g. cancers of the kidney, disorders of
 CC decreased phosphate levels including tumoral calcinosis, osteomalacia,
 CC osteoporosis, familial hypophosphataemia, rickets, cystinuria,
 CC nephrocalcinosis, glomerulonephritis, renal calculus, Alzheimer's
 CC disease, diabetes mellitus, hereditary amyloidosis, myopathies including
 CC progressive external ophthalmoplegia, Kearns-Sayre syndrome, myoclonic
 CC epilepsy, encephalopathy, cardiomyopathy, hypokalaemia, Goodpastures
 CC syndrome, and disorders of cell signalling through cAMP, ATP, NADPH and
 CC glucose-6-phosphate. Antagonists or inhibitors of NAPTR may be
 CC administered to a subject to treat or prevent disorders associated with
 CC increased phosphate levels, e.g. hypocalciuria, hypocalcaemia, and
 CC abnormal phosphate regulation in neurons, gastrointestinal tract and
 CC liver. The products can also be used for detection, diagnosis and drug
 CC screening.
 SQ Sequence 272 BP; 67 A; 56 C; 58 G; 91 T;

alignment_scores:
 Quality: 433.00 Length: 85
 Ratio: 5.094 Gaps: 0
 Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
 US-09-391-958-1 x V43712 ..

Align seg 1/1 to: V43712 from: 1 to: 272

```

317 11etyr11easvAl1euaSp11ealAProArGtyrSerSerPheLeuMe 333
|||||
1 ATTTATATCATGCTTATGATATTGCTCCAGGATATCCAGTTTTCAT 50
333 tG1yAlaSerArG1yPheSerSer11ealAProval111eValProthrv 350
|||||
51 GGGAGCATCAAGAGATTTCGAGCATAGACCTGTCATGTACCTGACCCTG 100
350 a1SerG1yPheLeuLeuSerG1nAsProG1uPheG1yTPAAsvAl 366
|||||
101 TCAGTGGATTTCCTTACGAGACCTCAGTTGGTGAGAGAAAGTC 150
367 PhePheLeuLeuPheAlaValAsnLeuLeuG1yLeuLeuPheYr11eU1 383
|||||
151 TTCTTCTGCTGTTGGCTTAACCTTTAGGACTACTCTTTCACCTCAT 200
383 epheG1yG1uAlaAsvAlG1ng1uTPAla1yG1uArG1ySLeuThra 400
|||||
201 ATTTGGAAGACAGATGTCCAGAAATGGCTAAAGAGAAATACTACTC 250
400 rG1eU 401
|||||
251 GTTTA 255

```

seq_name: N_Geneseq_36: V57926

seq_documentation_block:
 ID V57926 standard; DNA, 235033 BP.
 AC V57926;
 DT 23-DEC-1998 (first entry)
 DE Hereditary haemochromatosis subregion from an unaffected individual.
 KM Bovine butyrophillin; BF: human hereditary haemochromatosis; HFE;

KW diagnosis; Iron metabolism; NPT3; NPT4; RORet; BTF1; BTF2; BTF3;
 KW BTF4; BTF5; milk protein; lupus; Sjogren's syndrome; hypophosphataemia;
 KW type 1 sodium transport gene; ss.
 OS Homo sapiens.
 PN WO9814466-A1.
 PD 09-APR-1998.
 PF 30-SEP-1997: U17658.
 PR 07-MAY-1997: US-852495.
 PR 01-OCT-1996: US-724394.
 PA (PROG-) PROGENITOR INC.
 PI Feder JN, Kronmal GS, Lauer PM, Ruddy DA, Thomas WJ,
 PI Tsuchihashi Z, Wolff RK;
 DR WPI: 98-240014/21.
 PT Hereditary haemochromatosis gene products - used to develop products
 PT for the diagnosis and treatment of hereditary disorders in iron
 PT metabolism
 PS Example 2; Fig 8; 209pp; English.
 CC The present invention describes hereditary haemochromatosis gene
 CC products from the human haemochromatosis gene. The present sequence
 CC represents a hereditary haemochromatosis (Hh). Also described is a
 CC unaffected by hereditary haemochromatosis (Hh). Also described is a
 CC method to determine the presence or absence of the common hereditary
 CC haemochromatosis (HFE) gene mutation in an individual comprising:
 CC (a) providing DNA or RNA from the individual; and (b) assessing the
 CC DNA or RNA for the presence or absence of a haplotype or genotype where
 CC the presence or absence of the haplotype genotype indicates the likely
 CC presence of the HFE gene mutation in the genome of the individual. The
 CC HFE gene sequences from the present invention can be used to develop
 CC products for use in the diagnosis and treatment of HFE. The present
 CC invention also describes BTF genes, which are homologues of the milk
 CC protein butyrophillin (BTF), and can be used in the production of agonists
 CC and antagonists of BTF function. Also described are: (1) a RORet gene
 CC which can be used to develop products for the study, diagnosis and
 CC treatment of lupus and Sjogren's syndrome; and (2) NPT3 and NPT4 genes
 CC which are homologues of a type 1 sodium transport gene, and can
 CC similarly be used for hypophosphataemia.
 SQ Sequence 235033 BP; 68786 A; 48466 C; 49441 G; 68340 T;

alignment_scores:
 Quality: 318.00 Length: 543
 Ratio: 1.975 Gaps: 9
 Percent Similarity: 29.650 Percent Identity: 20.994

alignment_block:
 US-09-391-958-1 x V57926 ..

Align seg 1/1 to: V57926 from: 1 to: 235033

```

137 PherYr11ePhe.....G1yG1yAlG1yCysValCysCysLeuLeuTr 151
|||||
217123 TTTTATCCCTCTTCCAGGACATGGCTGTCTGCTGTCCTATG 217172
151 pPheValVal11eYrAsPAsPProvalSerTyPProTr11eSerThrs 168
|||||
217173 GTTCACAGTGTATTATGATGACCCCATGACATACCCGCGCTAAGTGTTA 217222
168 erG1u1yG1uYr11e11eSerSerLeu1yG1ng1uAlG1y..... 182
|||||
217223 GGGAAAGAGACACATCTCTGCTCAGCTGCTCAAGAGTACATGACACA 217272
183 .....Se 183
217273 CCTTGTACTGTGGCCCATGACAGAGTCTTAAAGGAGGAGGTGAGATCTC 217322
183 rSerTySGLn..Pro..... 187
217323 CTCTGAGAGGACCATCTTGGCTGCTCTAATATCATGATGATTAATCT 217372
188 .....LeuPro11e1yS1aMe1L 194
217373 TTCTTTTACGCCAGTTCTCTCTGACGAGCGTGTCCCATTAAGGCGATG 217422

```

```

194 euaGSerLeuProIleTTrpSerIleCysLeuGlyCysPheSerHisGln 210
    :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217423 TCACATGGCTACCACTTGGGCGCATTTCCGGGTTTTCAGGCATTTTC 217472
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
211 TrpLeuValSerThrMetValValTyrIleProThrIleSerSerVal 227
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217473 TGGTTGGCACATCATCTTCTTACATCTTACCAAGATATATACAGTACTCT 217522
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
227 lTyrHisValAsnIleArgAsp..... 234
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
217523 GCTCCATGTTTAAACATCAGAGATGTGATTTACTCTTACTACTTACGAA 217572
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217573 AATGATATGTGTAATTAAGAGAAACAGTCTGTGTACTTATACATTCT 217622
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217623 GCGTTTACATATAACCATTAATTTAACCTTCACATGACCTTGAGAGAG 217672
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217673 CATTGTTATATTCCTTTTACAGATGTGAAACAGACACTTAGAGT 217722
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217723 GAATATACCTTCCCCAGGTGTCACAACTACTAAGTAGAGCTGTGACG 217772
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217773 CATCATATATCTTAACCACTATGCTATACACACACACAGCTATTCGA 217822
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217823 AGCTTCTTTAGAAATTAATTTGCTGGCCAGGCATGTGGTCTATGCT 217872
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217873 GTAATTCGACCACTTGGGAGGCGCGAGGCAAGCATCATGAGTCAAGA 217922
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217923 ATGCAAGACCAAGCTGACCAATATGTTTACTAATATCATCTACTAAAA 217972
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
217973 AATCAAAAAATTAGCCAGGTGTGTGGCAGGCACTGTAAATCCAGCTATT 218022
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
218023 CAGGAGGCTGAGACAGAGAAATCGCTTGAAACCCAGAGGTGAGTTGCA 218072
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
218073 TTGAGCCAGATCATGACCTGACCTGAGCTGGGCGACAGAGTAAGAC 218122
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
218123 TCCGTTTCAAAAAACAAAAACCAAAATTAATTTGCTTTTATCTGGA 218172
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
218173 GCCCAGAGTGAAGCTTCTGGCCCTCTTATCTGAGACAGTCTCTTT 218222
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
218223 AGTGTGAAAAAGATGTAATTTTCCCAAAACACCAAGTATCATGG 218272
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
234 ..... 234
218273 GGGTAAGTAAATGGCTGCTGTGTACTGACAAATTTTGTGCTAACGT 218322
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 ..... AsnGlyLeuLeu 238

```

```

218323 ATCTCTATTACTACTCTGTATAACTTCCTTCCCTTCAGAGGAGTCTCG 218372
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
239 SerAlaLeuProPheIleValAlaTrpValIleGlyMetValGlyGly 255
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218373 TCCTCCCTGCTTTTATTTGCGTGCAGCAAGCTGTACATTTTAGAGGTCA 218422
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
255 rLeuAlaAspPheLeuLeuThrIleLys...PheArgLeuIleThrVal 271
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218423 GCTGGCAGATTTCTCTTTGTCAGGAATCTTTCACATGATCATCAGTGTG 218472
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
271 rGlyIleAlaIleThrIleLeuGlySer...Leu..... 280
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218473 GAAAGCTCTTTTATCTCTTGTGAAGATAGCGTGTGGCCCATTTAAC 218522
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
280 ..... 280
218523 CAATCCCTTTTTCGACATGTCTCAGAGGTTCCCGACACATGTCTC 218572
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
281 .....ProSer.SerAlaLeuIleValSerLeuProTy 291
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218573 CATTGCCAGGCGCTCCTCTCCATCAATATGTGCTGGCCCTGCCCTT 218622
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
291 rLeuAsnSerGlyTyrIleThrAlaThrAlaLeuLeuThrLeuSerCysG 308
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218623 TGTGGCCTCCAGTACTGATATACCATTTTGTCTGATACTATTTCG 218672
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
308 lLeuSerThrLeuCysGlnSerGlyIleTyrIleAsnValLeuAspIle 324
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218673 GGACCAAGTACCATATGTGACTCAGGTTTATCATCAACACCTTAGATATC 218722
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
325 AlaProArgTyrSerSerPheLeu 332
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
218723 GCCCCAGGTAAAGAGCTTACCTG 218746
    :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: N_Geneseq_36:V57903
seq_documentation_block:
ID V57903 standard; DNA; 237326 BP.
AC V57903;
DE 21-DEC-1998 (first entry)
DE Hereditary haemochromatosis subregion from an HH affected individual.
KW Bovine butyrophilin; BF; human hereditary haemochromatosis; HFE;
KW diagnosis; iron metabolism; NPT3; NPT4; hOret; BPF1; BPF2; BPF3;
KW BPF4; BPF5; milk protein; lupus; Sjogren's syndrome; hypophosphatemia;
KW type 1 sodium transport gene; ss.
OS Homo sapiens.
PN WO814466-A1.
PD 09-APR-1998.
PE 30-SEP-1997; U17658.
PR 07-MAY-1997; US-852495.
PR 01-OCT-1996; US-724394.
PA (PROG-) PROGENITOR INC.
PI Feder JN, Krommal GS, Laufer PM, Ruddy DA, Thomas WJ,
PI Tsuchihashi Z, Wolff RK;
DR WPI; 98-240014/21.
PT Hereditary haemochromatosis gene products - used to develop products
PT for the diagnosis and treatment of hereditary disorders in iron
PT metabolism
PS Claim 1; Fig 9; 209pp; English.
CC The present invention describes hereditary haemochromatosis gene
CC products from the human haemochromatosis gene. The present sequence
CC represents a hereditary haemochromatosis subregion from an hereditary
CC haemochromatosis (HH) affected individual. Also described is a
CC method to determine the presence or absence of the common hereditary
CC haemochromatosis (HFE) gene mutation in an individual comprising:
CC (a) providing DNA or RNA from the individual; and (b) assessing the
CC DNA or RNA for the presence or absence of a haplotype or genotype where
CC the presence or absence of the haplotype genotype indicates the likely
CC presence of the HFE gene mutation in the genome of the individual. The
CC HFE gene sequences from the present invention can be used to develop
CC products for use in the diagnosis and treatment of HFE. The present
CC invention also describes BPF genes, which are homologues of the milk

```


325 AlaProArgTyrSerSerPheLeu 332
 ||||| : : : : :
 220990 GCCCCAGGTAAGACTCTACTG 221013

seq_name: N_Geneseq_36:T62652

seq_documentation_block:

ID T62652 standard; cDNA to mRNA; 1840 BP.

AC T62652:

DE 15-MAY-1997 (first entry)

DE DNA encoding protein involved in cephalosporin C biosynthesis.

KW cephalosporin C; biosynthesis; fermentation; ds.

OS Acromonium chrysoeum.

FT Key location/Qualifiers

FT cds 127..1575

FT /*tag= a

PN J09009966-A.

PD 14-JAN-1997.

PE 03-JUL-1995; 167461.

PA (ASAH) ASAH KASEI KOGYO KK.

DR WPI: 97-126424/12.

PT Acromonium chrysoeum gene involved in biosynthesis of

PT cephalosporin C - used to improve fermentation ability of A.

PS Claim 3: Page 9-11; 21pp; Japanese.

CC This cDNA, derived from Acromonium chrysoeum, contains a coding

CC Sequence for a protein involved in biosynthesis of cephalosporin C.

CC The gene involved in biosynthesis of cephalosporin C or its cDNA

CC can be used to improve the fermentation ability of Acromonium

CC chrysoeum.

SQ Sequence 1840 BP; 346 A; 585 C; 446 G; 463 T;

alignment_scores:

Quality: 180.50 Length: 435

Ratio: 0.809 Gaps: 24

Percent Similarity: 51.264 Percent Identity: 23.678

alignment_block:

US-09-391-958-1 x T62652 ..

Align seg 1/1 to: T62652 from: 1 to: 1840

25 ValLeuHisPheCysAspThrThrIleAlaGlnAsnValIleMetAs 41
 : : : : :
 337 ATGTCGACTTTGACGACGACGACGACGACGACGACGACGACGACG 386
 41 nIle.ThrMetValAlaMetValAsn..... 49
 : : : : :
 387 TGTGTACATGTGCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCT 436
 50SerThrSerProGlnSerGlnLeuAsnAspSerSe 61
 : : : : :
 437 CCGAGATGTACGCGCGCGCTGACCTTACCTACCTACCTACCTACCT 486
 61 rGlValLeuProValAspSerPheGlyIleuSerIysAlaProLys 78
 : : : : :
 487 GTTCCTCTTACCA.....TCGGGTGCGCGCGCGCGCGCGCGCGCA 524
 78 eLeuProAlaIysSerSerIleLeuGly.....GlyGlnPheAlaI 92
 : : : : :
 525 CCGCGCGCGCTGTCATCTTCCGTATGCTGCGCGCGCATGAGGCTTC 574
 92 eTrrpGluArgTrrp.GlyPro.....ProGlnGlu 101
 : : : : :
 575 TGTATGACCATGTGGCGTGGCCACCGTGGCGGATCTTTTATCCAGA 624
 102 ArgSerArgLeuCysSerIleAlaLeuSerGlyMetLeuLeuGlyCys 118
 : : : : :
 625 CCGCGGTGGCGCGCATGCGCATTTGGACGTTTGGCGCGCGGTGATG 674

118 eThAlaIleLeuIleGlyIlePheIleSerGlnThrLeuGlyTrrpProp 135
 : : : : :
 675 TGTGTGCGCGCTGTGGAGGTATCTTGGCAGGAGGTGGTGGGT 724
 135 heValPheThrIlePheGlyIleValGlyCysValCysCysLeuLeuTrrp 151
 : : : : :
 725 GGGTGTGTTGGGTGTCGCGCATCGGGGGGGCTTCATACAGCATGTC 774
 152 PheValValIleTyrAspAspProValSerTyrProTrrpIleSerThr 168
 : : : : :
 775 TTCCTATACCCCGTAG.....ACCTATCCTCCGTTCTCTTAC 815
 168 rGluLysGluTrrpIleIleSerSerLeuLysGlnGlnValGlySerSer 185
 : : : : :
 816 GCGCAAG.....GTGAACGCTTACGACAAAGACCGCAAT.... 852
 185 ysgLInProLeu..... 188
 : : : : :
 853CCCCCTTACCTCAGCAGTACGCGACACCTCGAGTCAGCGCT 897
 189ProIleLysAlaMetLeuArgSerIle 197
 : : : : :
 898 CGGATTTCCCGGTGTGTCGCGCGCGCGCGCTCTCTCTCTCTCTCA 945
 197 uProIle.....TrrpSerIleCysLeuGly.....CysPheSerHisG 210
 : : : : :
 946 .CCCATCGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 994
 210 InTrrpLeuValSerThrMetValValTrrpIleProThrTrrpIleSer 226
 : : : : :
 995 AGTTCCTG.....CTTTGTTCATATCCCGAGTGTCTTGGCGAG 1035
 227 ValTrrHisValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuPro 243
 : : : : :
 1036 ATCTAGACTTCTCGCTGTGTCMAATAGG...CTCTCTATCTGGTAT 1082
 243 eIleValAlaTrrpValIleGly...MetValGlyGlyTrrpLeuAla 259
 : : : : :
 1083 TGGCGGCGCGCTGTGCTGTGGGAATGCATATTCGCGCAAGCCTGACC 1132
 259 heLeuLeuThrIleLys.....PheArg 266
 : : : : :
 1133 GATTCGTGCCAAAAGTCTGCGCGGATGAGTGAAGCGGAGTTTCGT 1182
 267 LeuIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSer 283
 : : : : :
 1183 CTTCCTCTGATG.....ATCCCC...GG 1202
 283 rAlaLeuIleValSerLeuProTrrpLeuAsnSerGlyTrrpIleThr 298
 : : : : :
 1203 GGCCTTTTGCATCCCATGTTTCTTCACTTACGCTGTGGGACCTTACT 1252
 299AlaThrAlaLeuLeuThrLeu 305
 : : : : :
 1253 ACAAGCTCACAGTGAATGATGCCATCTGCGCGCGCTCTCGGCAT 1302
 306 SerCysGlyLeuSerThrLeuCysGlnSerGlyIleTrrpIleAsnVal 322
 : : : : :
 1303 GGCCTTCAACCTGACTCATGAC.....ATACAGGCTTACTCTGT 1343
 322 uAspIleAlaProArgTrrpSerSerPheLeuMetGlyAlaSerArgGly 339
 : : : : :
 1344 AGACACCTACACACTGTACTCAGCTCGGCAATGGCGCGCGCACCAT 1393
 339 heSerSerIleAlaProValIleValProThrValSerGlyPheLeuLeu 355
 : : : : :
 1394 TCGGCTCTGTGGTGGTCTTTCCTCC...CTACCGCGCGCGCTCTCA 1440
 356 SerGlnAspProGluPheGlyTrrpArgAsnValPhePheLeuLeuPheAl 372
 : : : : :
 1441 TACGACGCGCTGAGTCTCGCTGGGGAATTCGACCTTGTGATTCATTGC 1490

```
372 aValAsnLeuGlyLeuLeuPheTyrlLeuIle...PheGlyGlyAla 387
|||||
1491 CGTGGCCCTCATACCTGTGCCTTCCATTATCATCCGCTATGGAGAGGGG 1539
seq_name: N_Geneseq_36:762653

seq_documentation_block:
ID 762653 standard; DNA; 3217 BP.
AC 762653;
DE 15-MAY-1997 (first entry)
DE DNA encoding protein involved in cephalosporin C biosynthesis.
KW cephalosporin C; biosynthesis; fermentation; ds.
OS Acromonium chrysogenum.
FH key location/Qualifiers
FT exon 935..1101
FT /tag- a
FT /note- "ATG start codon commences at nucleotide 935,
FT actual exon start not indicated in specification"
FT Intron 1102..1157
FT /tag- b
FT exon 1158..1614
FT /tag- c
FT Intron 1615..1675
FT /tag- d
FT exon 1676..2245
FT /tag- e
FT Intron 2246..2302
FT /tag- f
FT exon 2303..2602
FT /tag- g
FT /note- "actual exon end not indicated in specification,
FT nucleotide indicated here is end of coding
FT sequence, TGA stop codon found at 2603-2605"
PN 309009966-A.
PD 14-JAN-1997.
PE 03-JUL-1995; 167461.
PR 03-JUL-1995; JP-167461.
PA (ASAH ) ASAH KASEI KOGYO KK.
WP1 97-126424/12.
P-PSDB; M14439.
DR Acromonium chrysogenum gene involved in biosynthesis of
PT cephalosporin C - used to improve fermentation ability of A.
PT chrysogenum
PS Claim 4: Page 11-14; 21pp; Japanese.
CC This DNA, derived from Acromonium chrysogenum, contains a coding
CC sequence for a protein involved in biosynthesis of cephalosporin C.
CC The gene involved in biosynthesis of cephalosporin C or its cDNA
CC can be used to improve the fermentation ability of Acromonium
CC chrysogenum.
SO Sequence 3217 BP; 683 A; 934 C; 769 G; 798 T; 1 U;

alignment_scores:
Quality: 169.50 Length: 430
Ratio: 0.807 Gaps: 24
Percent Similarity: 48.837 Percent Identity: 23.953

alignment_block:
US-09-391-958-1 x 762653 ..

Align seg 1/1 to: 762653 from: 1 to: 3217

50 SerThrSerProGlnSerGlnLeuAsnAspSerSerGlyValLeuProVa 66
|||||
1364 GCCATCATCCCTACCATGTCACCAACGTCCTTCCTCCTTCACCA.. 1411
66 lAspSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLys 83
|||||
1412 .....TCGGGTGGCCCTCAGTCCCAACCTCCCGCCCTTGTCAT 1451
83 eSerIleLeuGly.....GlyGlnPheAlaIleTrpGluArgTrp. 97
|||||
1452 CTTCCGATGTGCTGGCCGCGCATGGAGGCTTCTGCGGTGATGACATGGCG 1501
```

```
97 lYPro.....ProGlnGluArgSerArgLeu 106
|||||
1502 GTGCCACCGTCGCCGATCTTTTATCCAGAGAGCGCGCTCGGCATG 1551
107 SerIleAlaLeuSerGlyMetLeuLeuGlyCysPheThrAlaIleLeu 123
|||||
1552 GCCATTGGAGCTTTGGCCCGTTGATGGGTCTCGTGTGGTCCGCGT 1601
123 eGlyGlyPheIleSerGlyThrLeuGlyTrpProPheValPheTyrlle 139
|||||
1602 GGGAGGATATCTTGGCCAGGCCAGGCTTGGCGGTGGGTGGTGGG 1651
140 .....PheGlyGlyValGlyCysValCysLeuLeuTrpPheValVal 154
|||||
1652 TCGCCATCGGGGTACGTAGTGTGTTCATTGTGAGATCGTTCTCCAG 1701
155 lIleTyraAsp..... 158
|||||
1702 GTTTCAGCTGACCGCATGTCTAGGGGGCTTCATAACAGCATGTTCTCC 1751
159 .....Pro.ValSerTyrrProTrpIleSerThrSerGlyLysGluTrp 172
|||||
1752 TCATAGCCCGTGAGACATATCCTCCGTTTCCTACAGGCAAG..... 1795
173 lIleIleSerSerLeuLysGlnGlnValGlySerSerLysGlnProLeu.. 188
|||||
1796 ..GTAAACCGCTACGACAGAAACCGGCAT.....CCCTCCT 1833
188 ..... 188
1834 TACCTACACACTAGCGGACCTCGATCGAGCGCTCGGATTTCCCGGT 1883
189 .....ProIleLysAlaMetLeuArgSerLeuProIle..... 199
|||||
1884 CTGTCGCCGCGCCCTCGTCTCTCTCCGCTCA...CCATCGCTGTC 1930
200 ...TrpSerIleCysLeuGly...CysPheSerHisGlnTrpLeuValSe 214
|||||
1931 CTCTCTCGCTTCATCGCGCTCGTCTTCATACGAGTTCTCTG..... 1975
214 TrpMetValValTyrlleProThrTyrlleSerSerValTyrlHisVal 231
|||||
1976 ...CTCTTGTGCACTATCCCAAGTGTCTTGGCAGATCTACGACTCT 2021
231 snIleArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrp 247
|||||
2022 CGCTTGTCATATAAGC...CTCTCTATCTGGGATATGGCGGCGCTG 2068
248 ValIleGly...MetValGlyGlyTyrlleAlaAspPheLeuThrLys 263
|||||
2069 CTGCTGGGGAATGCGCATATCGGCCAGCCTCTGACCGGATCTGCCA 2118
263 sLys.....PheArgLeuIleThrVal 271
|||||
2119 AAAGTCTGCGGAGATGAGTTGAMCCGAGATTCTCTCTGATG. 2167
271 rGlyIleAlaThrIleLeuGlySerLeuProSerSerAlaLeuIleVal 287
|||||
2168 .....ATCCC...GGGCTTTTGGCATC 2188
288 SerLeuProTyrlleAsnSerGlyTyrlleThr..... 298
|||||
2189 CCCATGTGTTCTTCATCTACGGCTGGCGACCTACTCAAGCTCCACTG 2238
299 .....AlaThrAlaLeuLeuThrLeuSerGlyLeu 310
|||||
2239 GATGATGCCCATCTGTGCGAGTGTCTCTCGGATTTGGCTCAACCTGT 2288
310 eT..... 310
|||||
2289 CACTCGTAAAGTCTTCTCTGCACTGGGATTTCTGAACAACCGTCTCAC 2338
```


FT	Key	Location/Qualifiers
FT	misc-feature	901..960
FT		/tag-a
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	2701..2760
FT		/tag-b
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	4501..4560
FT		/tag-c
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	6301..6360
FT		/tag-d
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	8101..8160
FT		/tag-e
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	9901..9960
FT		/tag-f
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	11701..11760
FT		/tag-g
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	13501..13560
FT		/tag-h
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	15301..15360
FT		/tag-i
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	17101..17160
FT		/tag-j
FT		/note="these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	misc-feature	
PN	EP-786519-A2.	
PD	30-JUL-1997.	
PE	07-JAN-1997.	100117.
PR	05-JAN-1996.	US-009861.
PA	(HUMA-) HUMAN GENOME SCI INC.	
PI	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA, Rosen CA.	
PI	WPI: 97-374922/35.	
DR	Polynucleotide(s) and proteins derived from <i>Staphylococcus aureus</i>	
PT	stored on computer readable medium and used in the production of	
PT	anti-S.aureus vaccines	
PS	Claim 1; Page 641-651; 3271pp; English.	

This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S. aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S. aureus infection. The polypeptides can also be used in a kit for the immunodetection of S. aureus in a sample. S. aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S. aureus DNA sequences contained on the computer readable medium.

Sequence 18613 bp; 5990 A; 2753 C; 3477 G; 5782 T;

alignment_scores:

	Quality:	138.50	Length:	392
	Ratio:	0.774	Gaps:	19.
	Percent Similarity:	45.663	Percent Identity:	20.663

alignment_block:
US-09-391-958-1 x V74423 ..

Align seg 1/1 to: V74423 from: 1 to: 18613

84 SerileLeuGlyGlnPheAlaIleTrrp..... 93
|||||:||||| :|||
9146 TCTGTAAATGGATTATTAATTCCTACTTTGGGACCTTAACGGGGTGTTCCA 9195
94 GluArgTrpGlyProp 99
9196 ATCAGTGTGGTGACCTGCAGATTATTCACAGATTTCAGAGTGGCGCCAA 9245
99 roGIngluarjserArjleucysSerlleAlaleuSerGlyMetleuLenu 115
9246 GAACGAACAGT..... GGCCGATACTTA 9268
116 GlycysPhe...ThralalleuIleGlylphenIleserGIurhle 131
||| :|||: ||||| | |||: :
9269 GATTTCGTGAATACATCACATAATATCGGTGTCGCATGACGAGTGGTGT 9318
131 uGly..TrpProPheValPheTrrlePheGlyGlyVal..... 143
::: ||| :|||: |||: |||
9319 TCCACTTGGGGTCTAATGTATTCCTCCATGAGAATATGTATTAGGATGT 9368
144 GlycysValCysCysleuLeutrpPheValIleIleTyr 156
9369 TCATTTCCCATCGGTGATGTCATTACTTATVTGGATCGACAATTAATTT 9418
157 AspAspProValserTyrProTrrPileserThrserGlu.. 169
||| ||||| :: ||| :|||: |||
9419 ATCGGAAAAGATGCCGAGAATAGAGATCCTGCTGAAGAAT 9468
170 LysGluTrrIleIleserSerleuLysGlnG 180
||| ||||| ||| |||: |||
9469 TTGGGAAGACCCTGCATTAAGAAAAATATGATTCCTCAA..... 9508
180 InvalIGlyserSerLysGlnProleuProIleLysAlaMetleuArgSer 196
9509 GGTrtGACGAATGGAGATCTTTAAAAAATATATATCTCTGGAAAT 9553
197 LeuProIleTrrPserIleCysleuGlyCysPheSerHisGlnTrrLeuVa 213
||||| :|||: ||| :|||
9554 CCTGTTATATGATTCATATG..... GT 9576
213 IserThrMetValValTrrIle..... P 221
|||||: ||| ||||| |

```

9577 TTCAAACGCTTTGTATACATGTACGAATCGATTTGATTAAGTGGGAC 9626
221 rorhryrileSerSerValTyrlhlsValaIenIeArgAspAsnGlyLeu 237
    |||:|||||:|
9627 CGTATATAGTGTACAG.....CATTTACCTTAGTAAAGGCGATGCA 9670
238 leuSerAlaIeuProPhe.....lleValAlaTrpValIleGlyMetVa 252
    :|:|:|:|:|:|
9671 GTTAAATACGATATCTTACTTTGAATATGTCATTAAGTTCAGATTTAT 9720
252 lGlylTyrlleuAlaIlePheleuLeuThrlYlYlPheArgleuIet 269
    :|:|:|:|:|:|
9721 ATGGGCGTACGATCAGAC.....TTATTTAA 9746
269 hrValAlaGlylIleAlaThrlleuGlySerleuProSerSerAlaIeu 285
    |||:|:|:|:|
9747 AAGGTGCGTGCNAATGTAGCTATTGGCTGATNGTTAATGATTAATT 9796
286 lIeValSerleuProTyrlleuAsnSerGlyTyrlleThralaThrala 302
    :|:|:|:|:|
9797 GTTGTCTTATTC.....TACCAATAGCTCAAGTGT 9828
302 uleuThrlleuSerGlyleuSerThrlleuGlySglnSerGlyIleTyrl 319
    :|:|:|:|:|
9829 CATGATGCTTAACATTTCAATTGTTGCATTA.....TTATTTAA 9859
319 leAsnValleuAspIleAlaProArg..... 327
9860 ..GTTGGGTAACTTTGTCGCCAATTATTAATGTTGATCANNNNNN 9907
328 .....Ty 328
9908 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 9957
328 rSerSerPheleuMetGlyAlaSerArgIlePheSerSerIleAlaPro 345
    :|:|:|:|:|
9958 NNNCGGCTATCTATTCGCTGACTCAATGGCGAAAGTTGGTGGCGCTA 10007
345 alIleValProthr.....ValSerGlyPheleuLeuSer 356
    :|:|:|:|:|
10008 TTGCTGATCCAAACGCTTAACGCTTTGATGATATACATTAAGT 10057
357 GlnAspProGlnPheGlyTrpArgAsnValPhePheleuPheAlaVa 373
    |||:|:|:|
10058 .....GGATGGACAGATGTTTCATGCTCTTATATGTC 10092
373 lAsnleuLeuGlyleuLeuPheTyrlleuPheGlyGluAlaAspValG 390
    :|:|:|:|:|
10093 ATTATTTCTTAGGCATG.....ATTCTATTAGAAATCGTTGCTTCT 10133
390 lnglUtrPalalysGluArgGlyLeu 398
    |||:|:|:|
10134 ATGAGAAAAAATAAGAGTTTA 10159
seq_name: N_Geneseq_36:V52165
seq_documentation_block:
ID V52165 standard; DNA; 10240 BP.
AC V52165;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:32.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN W09818931-A2.
PD 07-MAY-1998.
PE 30-OCT-1997; U19588.
PF 31-OCT-1996; US-023960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PT Kunsch CA, Rosen CA;
DR WPI; 98-272225/24.
Computer-readable medium with recorded Streptococcus pneumoniae

```

PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PS Claim 1: Page 326-332; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment of a sequence at least 95% identical
CC to SEQ ID NO:1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S. pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridize to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 10240 BP; 2865 A; 1914 C; 2390 G; 3068 T;

alignment_scores:

Quality:	131.00	Length:	336
Ratio:	0.834	Gaps:	12
Percent Similarity:	46.726	Percent Identity:	20.238

alignment_block:

US-09-391-958-1 x V52165 ..

Align seg 1/1 to: V52165 from: 1 to: 10240

```

68 SerPheGlyGlyleuSerIysAlaProIys..... 77
9218 ACTATGAGGAGGCTTGCGCTTGCCCAATATCTATGTTATCTTCT 9267
78 .....SerleuProAlaIysSerSerI 85
9268 TCGTTACTAAACGGGCTATTGACAGTTTGTCTTAATGCAACGGCAC 9317
85 lIeLeuGlyGlylnPheAlaIleTrpGluArgTrpGlyProProGlnGlu 101
    :|:|:|:|:|
9318 TGATAGCCAGTCAGCT.....CCAAAGGAG 9343
102 ArgSerArgleuGlySerIleAlaLeuSer...GlyMetleuLeuGlyCy 117
    :|:|:|:|:|
9344 AATACAGGCTCTGCTTAGGTACTTTGCTTACAGCGTAGTTGACGATAC 9393
117 sPheThralaIleleuIleGlylPheIleSerGlylnhrlleuGlyTrpP 134
    :|:|:|:|:|
9394 TCTAAGTGTCTTATTTAGGCTTTATGCGCAAAATATTATTTGGCATTC 9443
134 rOpheValPheTyrllePheGlyGlyValGlyCyValaCySglnleuLeu 150
    |||:|:|:|
9444 GTACAGTTTCTTACTGCTGTAGTATTCTATTATTAGCGCTATTTTG 9493
151 TrpPheValIleTyrlleAspAsp.....ProValSerTyrlProTrpI 165
    :|:|:|:|:|
9494 ACTATTTCTTATTCAGAGATTTTCACACAGTA..... 9529
165 eSerThrlSerGluTyrlleIleSerSerleuGlySglnGlnValG 182
9529 ..... 9529
182 lYSerSerIysGlnProleuProIleIysAlaMetleuArgSerleuPro 198
    :|:|:|:|:|

```

```

9530 . GCCAAGAAAAGGCTATTCCAAAGAAATATTACCGGTAA 9577
199 ILeTrrSerIleCysLeuGlyCysPheSerHisIleTrpLeuValSerTh 215
9578 TATCCCTATCTTTTGGTCAATCTCTT.....TTAACCAAGTTG 9615
215 rmetValValTyrIleProThrTyrIleSerSerValTyrHisValAsnI 232
9616 TGTCATCCAAATTTTCAGCTCAATCGATGGCCCAATTGGCTCTTATG 9665
232 leaArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpVal 248
9666 TAGCGGACTTAGGACACAGAGAATCTCTTTTGTCTCGGTGATTT 9715
249 ILeGlyMetValGlyGlyTyrLeuAlaAspPheLeuLeuThrIlySlySph 265
9716 GTGTCCGATGGGC..... 9730
265 eaTyrLeuIleThrValArgIlyIleAlaThrIleLeuGlySerLeuProS 282
9731 .....TTTCCAGCATGATGAGTGCAGAGTCATGCGCAAGCTAGCTG 9773
282 eISerAla.....LeuIleValSerLeuProTyrIleuAsn 293
9774 ACAAGTGGGCAATCATCGTCTGTGGTGTGCGCCAGTTTAT..... 9817
294 SerGlyTyrIleThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSe 310
9818 .....TCAGTCATCATCATCTCTCTGTCGCAATGCCCTC 9852
310 rThrLeuCysGlnSerGlyIleTyrIleAsnValLeuAspIleLeuAlaPro 327
9853 TAGCCCCCTTCAACTAGGACTTAT..... 9877
327 rGlyTrSerSerPheLeuMetGlyAlaSerArgIlyPheSerSerIleAla 343
9878 .....CGTTTCCTCTTGGATGGGAACCGT..... 9904
344 ProValIleValProThrValSerGlyPheLeuLeuSerGlnAspProG1 360
9905 ...GCCGTGATTCGCCGGGTAAAGCCCTACTCAGCAAAATGATCCCAA 9951
360 u.....PheGlyTrpArgAsnValPhePheLeuLeu 371
9952 AGCCGGCATTTCCAGGCTTTTGCCTTCATACAGTATCTTTATCTGG 10001
371 heAlaVal 373
10002 GAGGTGTT 10009
seq_name: N_Geneseq_36:V74708
seq_documentation_block:
ID V74708 standard; DNA; 415 BP.
AC V74708;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #397.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PD 30-JUL-1997.
PR 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines

```

```

PS Claim 1: Page 1292; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 415 BP; 122 A; 60 C; 88 G; 141 T;

alignment_scores:
Quality: 124.00 Length: 81
Ratio: 2.385 Gaps: 0
Percent Similarity: 64.198 Percent Identity: 30.864

alignment_block:
US-09-391-958-1 x V74708 ..

Align seg 1/1 to: V74708 from: 1 to: 415

96 TrrpGlyProProGlnArgSerArgLeuCysSerIleAlaLeuSerG1 112
112 TGGTTCCTTAAATAAGTAAAGTAGACATCAAGTCATATTATTAACAGG 205
112 ymetLeuLeuGlyCysPheThrAlaIleLeuIleGlyGlyPheIleSerG 129
206 ATCATATTGCGACCTGATTATGACCAATGATGATTCCTTTGTTA 255
129 IuThrLeuGlyTrrpProPheValPheTyrIlePheGlyGlyValG1Cys 145
256 ACGCATTTACTGCGACAGCACTATTATTCATTTTGGTCGAGTAGATTT 305
146 ValCysCysLeuLeuTrrpPheValValIleTyrAspAspProValSerT 162
306 TTAATGCGCKGATATATGGCGATATTATGCCAAGACTTACTGAGCRACA 355
162 rProTrrpIleSerThrSerGlyIlyGlyTyrIleIleSerSer 176
356 TAGMATGTTATATGACGCGAGAAACGTTTCATTATGAGAAAT 398
seq_name: N_Geneseq_36:X07357
seq_documentation_block:
ID X07357 standard; DNA; 1386 BP.
AC X07357;
DT 21-MAY-1999 (first entry)
DE Staphylococcus aureus tetracycline resistance gene.
KW Tetracycline resistance; antibiotic; antibacterial; screening;
KW infection; otitis media; endocarditis; abscess; conjunctivitis;
KW epididymitis; toxic shock syndrome; sepsis; septic arthritis;
KW diagnosis; therapy; vaccine; ds.
OS Staphylococcus aureus.
PN EP-893499-A2.
PD 27-JAN-1999.
PF 10-JUL-1998; 305495.
PF 23-JUL-1997; US-898976.

```

PA (SMK) SMITHKLINE BEECHAM CORP.
 PA (SMK) SMITHKLINE BEECHAM PLC.
 PI Burnham MKR, Lonetto MA, Warren PV,
 DR WPI: 99-097783/09.
 DR p-PSDB: W97820.
 PT New Staphylococcus aureus tetracycline resistant protein and gene
 PT useful as diagnostic reagents and for prevention and treatment of
 PT staphylococcal infections, including toxic shock syndrome and otitis
 PT media
 PS Claim 4: Page 25-27: 35pp: English.
 CC This is the coding region of a new Staphylococcus aureus gene
 CC encoding tetracycline resistance protein (see W97820) that is
 CC involved in sugar or drug transport across bacterial cytoplasmic
 CC membranes. The gene can be isolated from an S. aureus MCUH29
 CC (NCIMB 40771) DNA library. Vectors comprising the claimed
 CC polynucleotide, host cells and a process for producing the
 CC tetracycline resistance (TR) protein are claimed. TR polypeptides
 CC and polynucleotides are useful for diagnosing diseases related to
 CC over or under expression of TR protein by identifying mutations in
 CC the TR gene, or determining TR polypeptide or mRNA expression levels
 CC due to an infection of an organism with the TR gene (claimed).
 CC They can diagnose the stage and type of infection. TR polypeptides
 CC are also useful for screening for compounds which affect the
 CC activity. These can be used in treatment to inhibit (antagonist
 CC 1.e. antibacterial drugs) or enhance (agonist) TR activity. Direct
 CC administration of TR polypeptides can be used to treat conditions
 CC associated with a lack of TR (claimed). Direct administration of
 CC antisense oligonucleotides prevents expression. TR polypeptides
 CC (administered directly, in a vector (gene therapy) and as a vaccine)
 CC and antibodies induce an immune response to immunise and prevent
 CC disease (claimed). Diseases diagnosed, prevented or treated
 CC include: bacterial, especially S. aureus, infections of the
 CC respiratory tract (e.g. otitis media), cardiac (e.g. infective
 CC endocarditis), gastrointestinal (e.g. splenic abscess), CNS (e.g.
 CC cerebral abscess), eye (e.g. conjunctivitis), kidney and urinary
 CC tract (e.g. epididymitis and toxic shock syndrome), skin (e.g.
 CC wound infection), and bone and joint (e.g. septic arthritis). TR
 CC polypeptides, polynucleotides and their (ant)agonists can prevent
 CC adhesion of bacteria to matrix proteins, and are useful for use on
 CC wounds and body implants to prevent bacterial infection. TR is
 CC transcribed during both chronic and acute infections.
 SQ Sequence 1386 BP; 414 A; 212 C; 234 G; 526 T;

alignment_scores:
 Quality: 122.50 Length: 438
 Ratio: 0.583 Gaps: 20
 Percent Similarity: 47.945 Percent Identity: 18.265

alignment_block:

US-09-391-958-1 x X07357 ..

Align seg 1/1 to: X07357 from: 1 to: 1386

```

20 TYRGIYILEALALEUVALLEUHPHCEYSANPHETHTHRIEALAGL 36
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
274 TATGGCGGTATATTAATCAATT.....TTGCGATCAATTAATGG 311
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
36 nAsnValIleMetAsn.....IleThrMetValAlaMetValAsn 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
312 TAACATTTTCCAAATCAATCCCATTAACCTTAGTGGACGTATTATTC 361
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
50 eTThSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProVal 66
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
362 AAATGGCC..... 369
67 AspSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysSe 83
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
370 .....GGTTTACCTCTGCAGAGACA..... 390
83 rSerIleLeuGlyGlyGlnPheAlaIleTrp...GluArGTrpGlyProp 99
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
391 .....TTATATGTGATATATATCTGCCAAAGTATCTTTCTA 424

```

```

99 roGlnGluArSerArGlyLeuCysSerIleAlaLeuSerGlyMetLeuLeu 115
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
425 AAGGACGACGAGAACTTACCTTGCGTTAAGTACGACAGCACTATCTCTG 474
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
116 GlyCysPheThrAlaIleLeuIleGlyGlyPheIleSerGluTrpLeuG 132
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
475 TCATTAGTATATCGGTACATTATACAGGTGGATTAATTTTACGATTATTA 524
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
132 YTP.....Pro 135
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
525 CTGACAAATATATGTTTAAATGCAATTAACGATATTACGTGCCAT 574
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
135 heValPheTrpIle..... 139
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
575 TCATTATTAATTAATTAACCAAAAGAAATTAACGAATTAAGCTCATTTA 624
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
140 ...PheGlyGlyValGlyCysValCysLeuLeuTrpPheValValI 155
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
625 GATTTGTTGGCTTAATTTCTAGTGCACATTTGCTACACAGCTCATGCT 674
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
155 eTyrAspAspProValSerTyrProTrpIleSerTrpSerGluLysGlu 172
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
675 GTTT.....ATPACGAACTTAATGTTA.....T 700
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
172 YrIleIleSerSerLeu.....LysGlnGlnValGly 182
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
701 ATATGATTTGCTGCTTAATTAATGCAATTAATGCTTTTGCCTATATATAA 750
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
183 SerSerLysGlnProLeuProIleLysAlaMetLeuArSerLeuProI 199
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
751 AATGCCAAGCTGCATTAATTAATTAATCAATTTTCCAAATAAATACGTTA 800
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
199 eTrpSerIleCysLeuGlyCysPheSerIleGlnTrpLeuValSerTrp 216
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
801 TCGCTATATTTTATTAATTAATGATTTGTAATGTAATGCAATTCATGGGT 850
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
216 eValValYrIleProTrpTrpIleSerSerValYrIleIleValAsnI 232
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
851 ATATTTTACGTTCCATTCATTAATGACAAATTAATATCATCTGCAACGA 900
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
233 ArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpValI 249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
901 GACACAAACATCACTGTTATTAGTACCGGTTATATAGTAGCAGTCATCT 950
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
249 eGlyMetValGlyGlyTyrIleAlaAspPheLeuLeuTrpLysPheA 266
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
951 TCGTGCAATTAAGTGTAATTAATGCGCAATATCTGAATCAAAACAAGCA 1000
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
266 rGluIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSer 282
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1001 TTATCAGACAAATTAATTAATAGCAGCAGCTTG...ATTTTACCTGCA 1047
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
283 .....SerAlaLeuIleValSerLeu..... 289
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1048 TTTCAGTAGTAAATACATTTCAATCTTCGTCATTCATGATATCTT 1097
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
290 .....ProTyrLeuAsnSerGlyTyrI 297
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1098 TGCAGGTACCTTGTCTTAATGATAGCACCTTACTTAACGAA..... 1140
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
297 leThrAlaThrAlaLeuLeuTrpLeuSerCysGlyLeuSerThrLeuCys 313
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1141 .....GCCATTAACAAATAGATCTTAATATATCAGAGGTGGCT 1179
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
314 GlnSerGlyIleTyrIleAsnValLeuAspIleAlaProArGlyTrpSer 330
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1180 ...ATTGCTTTTATTAATTAATTAATTAATGCGG..... 1212
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
330 rPheLeuMetGlyAlaSerArGlyPheSerSerIleAlaProValIleV 347
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1213 .....GTATCTGTAGTATATGCAATTCGTCGCTCTAATACG 1249
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

347 alProHvaISerGlyPhe.....LeuLeuSerGlnAspPro 359
:::|||||
1250 ATTTTAAGCATTAAATATTTCCAGCAATGATGCATTAAATTCA..... 1293
:
360 GluPheGlyTyrParGAsnValaPhePheLeuLeuPheAlaValaAsnLeula 376
::::|||
1294 CATTTCGGT.....ATTATTTTATTTATTTTTAGGTTTAAAGAGATATGT 1337
:
376 uclYrLeuLeuPheTyrrLeuIlePheGlyGluAlaAspValoInglutPra 393
:|||||
1338 CGGATTAGATTATTATTCGTCATC.....TTAATCGTGGA 1372
:
393 lalysGluArgLys 397
:|||||
1373 CACAACTCGAAAAA 1386

THIS PAGE BLANK (USPTO)


```
|||||
335 TAGACAAAATGTCATCATGACATCACCATGGTAGCCATGGTCAACAGC 384
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
385 ACAAGCCCTCAATGCCAGCTCAATGATTCCTGAGGCTGCTGCTTTGA 434
67 pSerPheGlyGlyLeuSerLysAlaProLysSerLeuProAlaLysers 84
435 CTCATTTGGGGGCTAGTAAGGCCCAAGAGTCTTCGCAAAAGTCTT 484
84 erlleuenglyglyglnphealalletrpqluargrtrpqlpypProGln 100
485 CAATACTGGGGGCTCACTTTGCAATTTGGGAAAGTGGGGCCCTCCACA 534
101 GluArgSerArgLeuGlySerLysAlaLeuSerGlyMetLeuLeuGly 117
535 GAACGAAGCAGACTCTGCAGCATTCCTTTATCAGGAATGTTACTGGAT 584
117 spherhralalleuulleglylphelieserGIuThrleuGlyTrp 134
585 CTTTACTGCCATCCTCATAGTGGCTTTCATTTAGTAACCCCTGGGTGG 634
134 roPheValPheTrillePheGlyGlyValGlyValGlyCysLeuLeu 150
635 CCTTTGCTTCTATATCTTTGGAGGTGGCTGCTGCTGCTGCTTCTC 684
151 TrpPheValValIleTyrAspAspProValSerTyrProTrpIleSer 167
685 TGGTTTGTGTGATTTATGATGACCCCGTTTCTTATCCATGGATTAAG 734
167 rSerGluLysGluTyrIleIleSerSerLeuLysGlnValGlySers 184
735 CTCAGAAAAAGATACATCATATCCCTTGAACAACAGAGTGGGTCTT 784
184 erlysglnProleuproIleLysAlaMetLeuArgSerLeuProIleTrp 200
785 CTAAGCAGGCTCTCCATCAAAAGATATCTCAGATCTCTACCATTTGG 834
201 SerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSerThrMet 217
835 TCCATATGTTTAGGGCTTTCAGCCATCAATGTTAGTTAGCACATGGT 884
217 lValIlyrIleProThrTyrIleSerSerValIlyrHisValAsnIle 234
885 TGTATACATACCACTTACATCAGCTGTGTATCATCATGTTTACATCAG 934
234 sPasnglyLeuLeuSerAlaLeuProPheIleValAlaTrpValIleGly 250
935 ACAATGGACTTCTATCTGCCCTTCTTTATTTGTTGCTGGGTCAATAG 984
251 MetValGlyGlyTyrLeuAlaAspPheLeuThrLysLysPheArgle 267
985 AAGGGGGGGGCTATCTGAGATTTCTTCTAACCAAAAGTTAGACT 1034
267 ulleThrValArgLysIleAlaThrIleleuGlySerLeuProSerSera 284
1035 CATCACTGTGGGAAATGGCACAAATTTTGAAGAGTCTCCCTCTTAC 1084
284 lAleuIleValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThr 300
1085 CACTCATTTGTCTGTGCTTACCTCATTTCCGGGTATATACACCAACT 1134
301 AlAleuLeuThrLeuSerCysGlyLeuSerThrLeuGlyGlnSerGly 317
1135 GCGTTGCTGAGGCTCTCTGGGATTAAGACATGTGTCAAGTCAAGGAT 1184
317 eTyrIleAsnValleuAspIleAlaProArgTyrSerSerPheLeuMet 334
1185 TTAATCATATGCTTAGATATGCTCCAAAGTATTCAGTTTCTCATGG 1234
334 lYAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
|||||
```

```
1235 GACATCAAGAGATTTTTCAGCATAGCACCTGTCTATTTGACCACTGTC 1284
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTyrPArgAsnVal 367
1285 AGCGGATTTTCTTATGTCAGACCCCTGAGTTGGTGGAGGAAATCTTT 1334
367 ePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrIleu 384
1335 CTTCTTCTGCTTTGGCCGTTAACTGTTAGACTACTCTTCACTCATAT 1384
384 heGlyLysAlaAspValGlnGluTrpAlaLysGluArgLysLeuThrArg 400
1385 TTGGAGAACAGCATGTCTCCAAAGATGGGCTTAAGAGGAAACTCACTG 1434
401 Leu 401
1435 TTA 1437

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq.us-08-724-394A-19
seq_documentation_block:
; Sequence 19, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchinashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237/el
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOMSEND and TOMSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Flits, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 017957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ. ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1780 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1..1780
; OTHER INFORMATION: /note="cDNA 22E"
US-08-724-394A-19

alignment_scores:
Quality: 2061.00 Length: 401
```


Ratio: 5.165 Gaps: 0
Percent Similarity: 99.501 Percent Identity: 99.252

alignment_block:

US-09-391-958-1 x US-08-724-394A-19 ..

Align seg 1/1 to: US-08-724-394A-19 from: 1 to: 1780

```

1 MetGlnValAspGluThrLeuIleProArgLysValProSerLeuGlySse 17
|||||
377 ATGCAAGTGTAGAGACATGATCCCGAGGAAGGTCCAGATTATGTTTC 426
17 AlaArgTyrGlyIleAlaLeuValLeuHisPheCysAsnPheThrItr 34
|||||
427 TGCTCCGTAAGAAATAGCCCTCGTCTTACATTTCTGCAATTTCAACA 476
34 AlaGlnAsnValIleMetAsnIleThrMetValAlaMetValAsnSer 50
|||||
477 TAGCACAATAATGTCATCATGACATCCACATGGTAGCCATGGTCAACAG 526
51 ThrSerProGlnSerGlnLeuAsnAspSerSerGluValLeuProValAs 67
|||||
527 ACAAGCCCTCAATCCAGCTCAATGATTCCTGAGGTGCTGCTGCTTGA 576
67 PserPheGlyGlyLeuSerIysAlaProLysSerLeuProAlaLysSerS 84
|||||
577 CTGATTTGGTGGCCCTAAGTAAAGCCCAAGAGCTTCTGCAAAAGTCT 626
84 erIleLeuGlyGlyGlnPheAlaIleTrrPgluArgTrrPgluProGln 100
|||||
627 CAATACTGGGGGTGAGTTTGCAATTTGGGAAAGTGGGGCCCTCCACAA 676
101 GluArgSerArgLeuGlySerIleAlaLeuSerGlyMetLeuLeuGlyCy 117
|||||
677 GAACGAAGACACTGTCACGATTTGTTATCAGGAATGTTACTGGGATG 726
117 sPheThrAlaIleLeuIleGlyIlePheIleSerGluThrLeuGlyTrrP 134
|||||
727 CTTTACTGCCATCTCTCATAGTGGCTTTCATTAATGAAACCTTGGGTGC 776
134 robPheValPheTyrIlePheGlyIleValGlyCysValCysCysLeuLeu 150
|||||
777 CTTTGGTCTCTATATCTTTGGAGGTGGTGGTGGTGGTGGTGGTGGT 826
151 TrrPheValValIleTyrAspAspProValSerTyrProTrrPheSerTh 167
|||||
827 TGGTTTGTGTGATTTATGATGACCCCTTTTCTCATCCATGATGAACAC 876
167 rSerGluLysGluTyrIleIleSerSerLeuLysGlnGlnValGlySers 184
|||||
877 CTCAGAAAAGAAATACATCATATCTCTGAAACAACAGCGGGGTCTT 926
184 erLysGlnProLeuProIleLysAlaMetLeuArgSerLeuProIleTrrP 200
|||||
927 CTAAAGCAGCTCTTCCCATCAAAAGCTATGCTCAGATCTCACCATTTGG 976
201 SerIleCysLeuGlyCysPheSerHisGlnTrrPleuValSerThrMetVa 217
|||||
977 TCCATATGTTTAGCGTGTTCACGCCATCAATGATGTTAGTACACAAATG 1026
217 lValTyrIleProThrTyrIleSerSerValTyrHisValAsnIleArgA 234
|||||
1027 TGTATACATACCAACTTACATCAGCTCTGTGTACCATGTTAAACATCAGAG 1076
234 sPasnGlyLeuLeuSerAlaLeuProPheIleValAlaTrrPValIleGly 250
|||||
1077 ACAATGGACTTCTATCGCCCTTCTTATTTGTGGCTGGGTCAATAGGC 1126
251 MetValGlyGlyTyrLeuAlaAspPheLeuLeuThrLysLysPheArgLe 267
|||||
1127 ATGGTGGAGAGCTATCTGCGCAGATTTCTCTTAACCAAAAAGTTTAAAGCT 1176
267 uIleThrValArgLysIleAlaThrIleLeuGlySerLeuProSerSerA 284

```

```

|||||
1177 CATCAGCTGAGGAAATTCGCACAAATTTTAGAGAAAGTCTCCCTCTCAG 1226
284 lAleuIleValSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThr 300
|||||
1227 CACTCATTTGTGTCTCTCTTACTCTCAATTCCGGCTATATCACAGCAACT 1276
301 AlAleuLeuThrLeuSerCysGlyLeuSerThrLeuGlyGlnSerGlyI 317
|||||
1277 GCTTGTGACGCTCTCTTCGCGATTAAGCACATTTGTGATCAGGAGAT 1326
317 eTyrIleAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetG 334
|||||
1327 TTATATCAATGTCCTTATATATGCTCCAGGATATCCAGTTTCTCATGG 1376
334 lAlaSerArgGlyPheSerSerIleAlaProValIleValProThrVal 350
|||||
1377 GAGCATCAAGAGGATTTTCAGACATAGCACCTGCTCATTTGACCACATG 1426
351 SerGlyPheLeuLeuSerGlnAspProGluPheGlyTrrPArgAsnValPh 367
|||||
1427 AGCGGATTTCTTCTTACTCAGGACCCGAGATTGGGTGGAGGAATGCTCT 1476
367 ePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrIleThr 384
|||||
1477 CTTCCTCTGTTTGGCCCTTAACCTGTTAGACATCTTCTTACTCTCATAT 1526
384 heGlyGlnAlaAspValGlnGluTrrPAlaLysGluArgLysLeuThrArg 400
|||||
1527 TTGGAGAAGCAGATGTCACGAAGATGGGCTTAAGAGAGAAAAGTCACTCG 1576
401 Leu 401
|||
1577 TTA 1579

seq_name: /cgn2_6/plodata/1/lna/5C_COMB.seq:US-08-724-394A-18
seq_documentation_block:
; Sequence 18, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Therefo
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/724,394A
; FILING DATE: 01-OCT-1996
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Fitts, Renee A.
; REGISTRATION NUMBER: 35,136
; REFERENCE/DOCKET NUMBER: 01957-000100
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200

```

```

; TELERAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 2266 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: not relevant
;   TOPOLOGY: not relevant
;   MOLECULE TYPE: cDNA
;   FEATURE:
;     NAME/KEY: misc.feature
;     LOCATION: 1..2266
;     OTHER INFORMATION: /note= "cDNA 22B"
US-08-724-394A-18
```

```

alignment_scores:
    Quality: 932.00      Length: 471
    Ratio: 3.305        Gaps: 4
    Percent Similarity: 59.873    Percent Identity: 41.826
```

alignment_block:

US-09-391-958-1 x US-08-724-394A-18

Align seg 1/1 to: US-08-724-394A-18 from: 1 to: 2266

```

10 ArglyValProSerLeuGysSerAlaArgTylGlyIleAlaLeuValLe 26
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
440 AGGAAAGGTCAGAGATTCTGTTCATACGCTAATGGCTGGCTCTTATCAT 489
26 uHisPheCysAsnPheThrIleAlaIleAsnValIleMetAsnIleT 43
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
490 GCACCTTCACAACTCCACCATGATACGCGGTGTGAGCTGTGAGCATG 539
43 hMetValAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAp 59
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
540 CGATCATGCCATGGTGAACACACACACACACAGAGCTATCTAATGCC 589
60 SerSerGluValLeuProValAspSerPheGlyGlyLeuSerLysAlaBr 76
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
590 TCCACTGAGGGGCGTGTGACAGATGCTTCATTAACCTCAGCATATCCAT 639
76 oLysSerLeuProAlaLysSerSerIle..... 85
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
640 CAGGAATTTGATACAAAGGCTGTGTATCATCAATGAGCCAGAAACTC 689
85 ..... 85
690 AGGTATCATCTTTAGCTCCATCACTAATGAGTAACTGACTGATC 739
86 ..... 87
740 CCAGTGTATATTAGCAGGATATTGGAGCAAAAAAATGCTGTGTC 789
87 ..... 87
790 TGGTTGCTGATCTCTTCCCTTCTCACCCTCTTTACACCACCTGGCTG 839
87 ..... 87
840 ACTTCGAGATGATTTTGGTCATCATGTTGCGAGACGTCAGGGCATGCC 889
88 ..... 88
890 CAGGGAATGGCATGAGCAGGTGATGTTACTATTGGGCAAGTGGGCTCC 939
98 oProGlnIleuArgSerArgLeuGysSerIleAlaLeuSerGlyMetLeuL 115
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
940 TCCACTTCAAGGAAGCAGCTCACCACATTCGAGATCAGGGTCAGCAT 989
115 euGlyCysPheThrAlaIleLeuIleGlyGlyPheIleSerGluThrLeu 131
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
990 TTGGATCCTTCATCATCTCTGTGTGGGGGAGACATAATCTCAGAGCCCTT 1039
```

```

132 GlyTrpProPheValPheTrpIlePheGlyValGlyCysAlaCys 148
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1040 AGCTGGCCTTTATCTTACATCTTGTGAGCAGCTGGCTGTGCTG 1089
148 sLeuLeuTrpPheValValIleTrpAspAspProValSerTyrProTrpI 165
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1090 TCTCTATGAGTTCACAGTGTATGATGATGACCCCATGACATCACCCGTGA 1139
165 lSerThrSerGlyLysGlyTrpIleIleSerSerLeuLysGlnVal 181
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1140 TAAGTTAGGGGAAAGAGACATCTCTCCATCGCTCAACAGCC 1189
182 GlySerSerLysGlnProLeuProIleLysAlaMetLeuArgSerLeu 198
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1190 AGTTCTCTGGAGCAGCTGTCCCATAAAGCGAGTGTCAATGCTTACC 1239
198 oIleTrpSerIleCysLeuGlyCysPheSerHisGlnTrpLeuValSer 215
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1240 ACTTGGCCATTTTCTGCTGGTTTTCAGCCATTTCTGTGTATGACCA 1289
215 hMetValValTyrIleProThrTyrIleSerSerValTyrHisValAsn 231
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1290 TCATCCTACATACCTACCAACGATATATCATCTGCTCATGTATAC 1339
232 lLeuArgAspAsnGlyLeuLeuSerAlaLeuProPheIleValAlaTrpVa 248
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1340 ATCAGAGATAGTGGAGTGTGTCTCTCCCTGCTTTATTTGCTGCGCAG 1389
248 lIleGlyMetValGlyGlyTyrIleuAlaAspPheLeuThrLysLys 264
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1390 CTTGACAAATTTAGGAGGTGCACCTGCGACATTTCTTTGTGTACAGAAAC 1439
265 .. PheArgLeuIleThrValArgLysIleAlaThrIleLeuGlySerLeu 280
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1440 TTCTGAGATGTATCATCTGTGCGAAAGCTTTTTCATCTCTTCA..... 1482
281 ProSerSerAlaLeuIleValSerLeuProTyrLeuAsnSerGlyTyrI 297
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 ..... 1482
297 eThrAlaThrAlaLeuLeuThrLeuSerCysGlyLeuSerThrLeuGys 314
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1482 ..... 1482
314 lNserGlyIleTyrIleAsnValLeuAspIleAlaProArgTyrSerSer 330
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1483 ..... 1491
331 PheLeuMetGlyAlaSerArgLysPheSerSerIleAlaProValIleVa 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1492 TTCTCATGGGAATCTCAAGGGGATTTGGGCTCATCGCAGAAATCATCTC 1541
347 lProThrValSerGlyPheLeuLeuSerGlnAspProGluPheGlyTrpA 364
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1542 TTCCACTGCCACTGGATTCATCATGACGAGATTTGAGCTGGTGTGA 1591
364 rGAsnValPhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPhe 380
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1592 GGAATGTCTTTTCTGTCTGTCTGAGTCAACATGTTGGCTGGCTGCTT 1641
381 TyrLeuIlePheGlyGluAlaAspValGlnIleuTrpAlaLysGluArgLy 397
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1642 TACCTCAGTGTGGCAAGCAGAACTTCAAACTGAGGCGCAAGAGAGAC 1691
397 sLeuThrArgLeu 401
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
1692 CCTTACCGGCTC 1704
seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-647-484-1
seq_documentation_block:
; Sequence 1, Application US/08647484
; Patent No. 5618677
```


386 luAlaaspValgInGIuTPAlaIySgluaRgylseuThr 399
: : : : :
1932 CTGGAGAGAGACGCGCTGGCAGACCTGAGAGATGAC 1972

seq_name: /cgm2_6/ptodata/1/ina/5A_COMB.seq:US-08-647-484-3

seq_documentation_block:

Sequence 3, Application US/08647484
Patent No. 5618677
GENERAL INFORMATION:
APPLICANT: NI, Binhui
APPLICANT: Paul, Steven M.
TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/647,484
FILING DATE: 14-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/430,033
FILING DATE: 27-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Gaylo, Paul J.
REGISTRATION NUMBER: 36,808
REFERENCE/DOCKET NUMBER: X-10006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (317) 276-0756
TELEFAX: (317) 276-3861
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2716 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: mRNA
US-08-647-484-3

alignment_scores:

Quality: 479.00 Length: 397
Ratio: 2.056 Gaps: 7
Percent Similarity: 58.690 Percent Identity: 28.212

alignment_block:

US-09-391-958-1 x US-08-647-484-3 ..

Align seg 1/1 to: US-08-647-484-3 from: 1 to: 2716

12 ValProSerLeuCySerAlaTrgTyrGlyIleAlaIleuValLeuNisph 28
: : : : :
941 AUCCGCUACGUGCCGCGGUCACUAGGCGUGGUCACUC..... 979
28 eCysAsnPhetThrTrpIleAlaGlnAsnValIleMetAsnIleThrMetV 45
: : : : :
980UUCGUGAGAGAUCCGACAGGCGUGUGAGAGGCGUCACUAC 1022
45 aAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
: : : : :
1023 CCGCCUGC..... 1030

62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaIleProIysSe 78
1030 1030
78 IleuProAlaIysSerSerIleuGlyGlyIlePheAlaIleTrpGluA 95
1031CAUGGAGUUGGACGA 1046
95 rGTrpIleProProGlnGluArgSerArgLeuCySerIleAlaLeuSer 111
: : : : :
1047 AAUGGCGCCACCCUUGAAGAGGCGCCUGCGGACGACGACGCUUUUU 1096
112 GlyMetLeuLeuGlyCysPheThrAlaIleIleuIleGlyGlyPheIleSe 128
1097 GGUUCCUAGUGGUGGCGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1146
128 rGluThrIleuGlyTrpProPheValPheTrpIlePheGlyValGlyC 145
: : : : :
1147 GCAGUACUACAGAGAGGCGUUGUUCUACGUCUACGCGACGCUUGGGA 1196
145 ysValCysCysLeuLeuTrpPheValIleTrpAspProValSer 161
1197 UCUUCUGUACCUUGUUCUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1246
162 TyrProTrpIleSerThrSerGlyGlyTrpIleIleSerLeuLeu 178
: : : : :
1247 CACCCGACAUCCUGGAGAGGCGCAAGUACUC.....GA 1284
178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188
1285 GGCAGCGCAUCCGAGAGGCGCAAGUACUAGAACCCCGCAGAGUUA 1334
189ProIleLysAlaMetLeuArgSerLeuProIleTrpSerIleCys 203
1335 GCACUCCUGGCGGCGGCGUUCUACGUCUACGUCUACGUCUACGUC 1384
204 LeuGlyCysPheSerHisGlnTrpIleValSerThrMetValIleTrp 220
1385 GUGGCGACUUCUUGCGCGGAGGAGGUGUUCUACGUCUACGUCUAC 1434
220 eProThrTrpIleSerSerValIleTrpHisValAsnIleArgAspAsnGly 237
1435 GCGCGACUACUUCGAGAGAGGUGGUGGUGGUGGUGGUGGUGGUGG 1484
237 euleuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
1485 DGGUGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1534
254 GlyTyrLeuAlaAspPheLeuLeuTrpHisLysPheArgLeuIleThr 269
1535 GGCAGAUCCGCGGACUUCUUGCGGAGCGCGCGGACUAGUUCACACCA 1584
270 ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 286
1585 CGUGGCGAUGUUGAUGAUGGCGGAGGCGUUGCGCAUGAGAGCAGCUC 1634
286 IeValSerLeuProTyrIleuAsnSerGlyTyrIleThrAlaThrAlaLeu 302
1635 UGUUGUGGUGGCGUACUCCGACUCCAGGCGGUG...GCCAUUCUUC 1681
303 LeuThrIleuSerCysGlyLeuSerThrIleuGlyGlnSerGlyIleTyr 319
1682 CUGGUGCCUAGCGUGGUGGUGGUGGUGGUGGUGGUGGUGGUGGUG 1731
319 eAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAlas 336
1732 GAACACGUGGAGACUAGCCCGCGGCUAGCCGCAUUCUCCAGAGGACU 1781
336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly 352
1782 CCAAGCGGCGGCGGACACUUGCGGCGGAGGAGGUGGUGGUGGUGGUG 1831

```

seq_name: /cgn2_6/ptodata/1/lna/5A_COMB.seq:US-08-647-481-1
seq_documentation_block:
; Sequence 1, Application US/08647481
; Patent No. 5618918
; GENERAL INFORMATION:
; APPLICANT: NI, Binhui
; APPLICANT: Paul, Steven M.
; TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
; TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/647,481
; FILING DATE: 14-MAY-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/430,033
; FILING DATE: 27-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2716 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 461..2140
; US-08-647-481-1

Alignment_scores:
Quality: 479.00 Length: 397
Ratio: 2.056 Gaps: 7
Percent Similarity: 58.690 Percent Identity: 28.212

alignment_block:
US-09-391-958-1 x US-08-647-481-1 ..
Align seg 1/1 to: US-08-647-481-1 from: 1 to: 2716
12 ValProSerLeuGlySerAlaArgTyrGlyTleAlaLeuValLeuHisPn 28

```

[illegible]

```

1682 CTGCTCTACCGCTGGGCTTCAGCGGCTTCGCATCTCTGGTTCAACGT 1731
319 easnValleuAspIleAlaProArgTyrSerSerPheleuMetGlyAla 336
1732 GAACACCTTGACATAGCCCCGCTACGACGACATCTCATATGGCATCT 1781
336 eArGlyPheSerSerIleAlaProValIleAlaProThrValSerGly 352
1782 CCAGAGCGCGGGGACACCTGCGGCATGTGCCCCATCATCGGGG 1831
353 PheleuLeuSerGlnAspProGluPheGlyTyrPArgAsnValPhePhe 369
1832 GCCATACACACCAAGACTCCGGAGAGAGTGCAGTACGTCTTCTTAAT 1881
369 uleuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrIleIlePheGly 386
1882 TGCCTCCCTGGTGTCACATGAGAGGTGTCATCTTACGGGCTTTGGCTT 1931
386 lAlaAspValGlnGluTyrPalaIleGluArgLysLeuThr 399
1932 CTGAGAGAGAGACGCGCTGGCGACGCGTACGAGAGATGAGAC 1972

```

seq_name: /cgn2_6/ptodata/1/ina/5A-COMB.seq:US-08-647-481-3

```

seq_documentation_block:
: Sequence 3, Application US/08647481
: Patent No. 5618918
: GENERAL INFORMATION:
: APPLICANT: NI, Binhut
: APPLICANT: Paul, Steven M.
: TITLE OF INVENTION: HUMAN BRAIN SODIUM SODIUM DEPENDENT INORGANIC
: TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Eli Lilly and Company
: STREET: Lilly Corporate Center
: CITY: Indianapolis
: STATE: Indiana
: COUNTRY: United States of America
: ZIP: 46285
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/647,481
: FILING DATE: 14-MAY-1996
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/430,033
: FILING DATE: 27-APR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Gaylo, Paul J.
: REGISTRATION NUMBER: 36,808
: REFERENCE/DOCKET NUMBER: X-10006
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (317) 276-0756
: TELEFAX: (317) 276-3861
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2716 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: mRNA
: US-08-647-481-3

```

```

alignment_scores:
  Quality: 479.00      Length: 397
  Ratio: 2.056        Gaps: 7
  Percent Similarity: 58.690      Percent Identity: 28.212

```

alignment_block:
US-09-391-958-1 x US-08-647-481-3 ..
Align seg 1/1 to: US-08-647-481-3 from: 1 to: 2716

```

12 ValProSerLeuGlySerAlaArgTyrGlyIleAlaLeuValLeuHisPh 28
941 AUCCGCCACACUCCCGCGUCACUAUGGUGUGUUAUC..... 979
28 ecysAsnPheThrThrIleAlaGlnAsnValIleMetAsnIleThrMet 45
980 .....UUCUGAGAGAUCCUGCAGGGGUGUGAGAGGGGUGUACUACC 1022
45 AlAlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
1023 CCGCCUGC..... 1030
62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLysSe 78
1030 ..... 1030
78 rleuProAlaLysSerSerIleLeuGlyGlyGlnPheAlaIleTyrGluA 95
1031 .....CAUGGAGUCUGAGCA 1046
95 rGTrGlyProProGlnGluArgSerArgLeuGlySerIleAlaLeuSer 111
1047 AAUGGGCCCCACCCUUGAAGACGAGCUGCCUGGCGACAGACCUUUGU 1096
112 GlyMetLeuLeuGlyGlySerPheThrAlaIleIleGlyIlePheLe 128
1097 GUUUCUUAUGCGUGGGGUGUGUGCGAUGCCUUGCGGUGUCUUU 1146
128 rGluThrLeuGlyTyrProPheValPheTyrIlePheGlyIleValGly 145
1147 GCAGUACUACAGUAGAGCUUGUUUUCACGUCUACGCGACUUCGGA 1196
145 ysValGlyScysLeuLeuTyrPheValValIleTyrAspAspProVal 161
1197 UCUUCUGUACUUCUUCUGUGUCUUCUGUCUUCUACAGACCCCGGUG 1246
162 TyrProTyrIleSerThrSerGluGlySerGlyTyrIleIleSerSer 178
1247 CACCCAGCAUUCUGGAGAGAGCGCAAGUUAUC.....GA 1284
178 sGlnGlnValGlySerSerLysGln.....ProLeu..... 188
1285 GGACGCCAUCGGAGAGCGGCAAAUCUAGAACCCUACAGAAUGUUA 1334
189 .....ProIleLysAlaMetLeuArgSerLeuProIleTyrSerIle 203
1335 GCACUCCUGCGGCGGCGUUCUACGUCUACGUCUACGUCUACUAC 1384
204 LeuGlyCysPheSerHisGluTyrLeuValSerThrMetValValTyr 220
1385 GUGGCCACUUCUUGCCGCGACGUGGAGUGUUCUACUGUCUUAUC 1434
220 eProThrTyrIleSerSerValTyrHisValAsnIleArgAspAsnGly 237
1435 GCCCGACUACUUCGAAAGAUUGUUGGCUUGACAGUACGAAAGUAGGC 1484
237 uleuSerAlaLeuProPheIleValAlaTyrValIleGlyMetValGly 253
1485 UGUGUUCGCGCGCCGCCACUUGUACAUAGCAUACUUGGCCAUUG 1534
254 GlyTyrLeuAlaAspPheLeuLeuThrLysLysPheArgLeuIleThr 269
1535 GGCGAUGCGGACUUCUUGGAGCGCGCCGACUACUUGUACACCA 1584
270 ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu 286
1585 CGUGGCAAGUUGAUGAACUUGCGGAGGCUUGGCAUGAGACGACGUG 1634

```



```

254 GlyTyrLeuAlaaspheLeuThrLysLysPheArgLeuIleThr.. 269
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1535 GGCCAGATCGGGGACTCTCTCGGAGCGCGGATCATGTCACACACCAA 1584
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
270 ValArgLysIleAlaThrIleLeuGlySerLeuProSerSeraLeuI 286
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1585 CGTGGCGAGTGTATGATGACGTGGGAGGCTTCGGCATGAGACCGCT 1634
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
286 LevalSerLeuProTyrLeuAsnSerGlyTyrIleThrAlaThrAla 302
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1635 TGTTGGTGGTGGCTACTCGCATCCACAGGCGTG..GCCATCTCTTC 1681
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
303 LeuThrLeuSerCysGlyLeuSerThrLeuCysGlnSerGlyIleTyr 319
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1682 CTGTGCTTACCGCTGGGCTTCAGGGGCTTCGCCATCTCTGGGTTCAC 1731
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
319 eAsnValLeuAspIleAlaThrProArgTyrSerSerPheLeuMetG 336
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1732 GAACCACTGACATAGCCCCGGCTACGACGACATCTCATGGGCTATCT 1781
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerG 352
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1782 CCAAGCGCGTGGGACACACTCGGCGCATGTGTGCCCATCATCTGGGG 1831
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
353 PheLeuLeuSerGlnAspProGluPheGlyTyrPargAsnValPheP 369
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1832 GCCATGACTAGACACAGACTCGGAGAGTGCAGACAGCTGTCTCTTA 1881
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
369 uLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrIleIlePheG 386
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1882 TGCCCTCCCTGGTCACTATGAGAGGTGTCTTACGGGCTCTTGCTT 1931
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
386 lValAspValGlnGluTyrPalaIleGluArgLysLeuThr 399
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1932 CTGAGAGAGAGAGCGCGTGGCGAGCGTGAAGAGATGAGC 1972
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
seq_name: /cgn2_6/plodata/1/lna/5A_COMB.seq:US-08-430-033A-3
seq_documentation_block:
; Sequence 3, Application US/08430033A
; Patent No. 5686266
; GENERAL INFORMATION:
; APPLICANT: NI, Binhui
; APPLICANT: Paul, Steven M.
; TITLE OF INVENTION: HUMAN BRAIN SODIUM SODIUM DEPENDENT INORGANIC
; TITLE OF INVENTION: PHOSPHATE COTRANSPORTER AND RELATED NUCLEIC ACID COMPOUNDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Lilly Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/430.033A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Gaylo, Paul J.
; REGISTRATION NUMBER: 36,808
; REFERENCE/DOCKET NUMBER: X-10006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317) 276-0756
; TELEFAX: (317) 276-3861
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2716 base pairs

```

```

; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: mRNA
; US-08-430-033A-3

alignment_scores:
    quality: 479.00    length: 397
    ratio: 2.056      gaps: 7
    percent similarity: 58.690    percent identity: 28.212

alignment_block:
US-09-391-958-1 x US-08-430-033A-3 ..
Align seg 1/1 to: US-08-430-033A-3 from: 1 to: 2716

12 ValProSerLeuCysSerAlaArgTyrGlyIleAlaLeuValLeuHisph 28
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
941 AUCCCTCAGACUCCGCCGUCACUAGGCUUGUCAUC..... 979
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
28 eCysAsnPheThrThrIleAlaGlnAsnValIleMetAsnIleThrMet 45
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
980 .....UUCGUGAGAUCCUCCGACGGGUGUGUAGAGGGGUGUACAUACC 1022
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
45 AlaMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1023 CCGCCUCC..... 1030
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
62 GluValLeuProValAspSerPheGlyGlyLeuSerLysAlaProLys 78
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1030 ..... 1030
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
78 rLeuProAlaLysSerSerIleLeuGlyGlnPheAlaIleTyrGln 95
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1031 .....CAUGGAGAUCCGAGCA 1046
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
95 rGTrpGlyProProGlnGluArgSerArgLeuCysSerIleAlaLeuSer 111
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1047 AAUGGCCCCACCCUUGAAGAGAGUCCGUGGACGACGACACCCUUUUU 1096
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
112 GlyMetLeuLeuGlyCysPheThrAlaIleLeuIleGlyPheIle 128
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1097 GUUUCUUAUGUGGGGCGGUGUGUGGAGNCCGCCGCGGGUCCUUU 1146
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
128 rGluThrLeuGlyTyrProPheValPheTyrIlePheGlyValGly 145
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1147 GCAGUACUACGAGUAGGACUUGUUUUCUACGUCUACGCGACUUCGG 1196
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
145 yValCysCysLeuLeuTyrPheValIleTyrAspAspProValSer 161
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1197 UCUUCUGUACUUGUUCUGUGUGUCCUCCAGAUCCGCCCGCGUG 1246
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
162 TyrProTyrIleSerThrSerGluLysGluTyrIleIleSerSerLeu 178
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1247 CACCCAGACUUCUGGAGAGAGGCAAGUAC.....GA 1284
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
178 sGlnIleValGlySerSerLysGln.....ProLeu..... 188
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1285 GAGCCGCAUCCGAGAGAGCGGCAACUACUAGAACCCUCCAGAGUUUA 1334
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
189 .....ProIleLysAlaMetLeuArgSerLeuProIleTyrSerIle 203
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1335 GCACUCCUCCGCGGCGGCUUCUACGUCUUAAGCCAGUUAAGCCAUCC 1384
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
204 LeuGlyCysPheSerHisGlnTyrLeuValSerThrMetValIleTyr 220
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1385 GUGGCCAACUUCUCCGACGUCGACGUCUACUCCGUCUACUCCCA 1434
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
220 eProThrTyrIleSerSerValTyrHisValAsnIleArgAspAsnGly 237
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1435 GCCCGACAUCCUAGAAAGUUGUGGCUUCGAGAUCCAGCAAGUAGGCC 1484
    ||| :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```



```

237  enLeuSerLheAlaAspPheLeuLeuThrIleValAlaTrpValIleGlyIleMetValGly  253
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1485  DeGUUGCCGGCGGCGGCCACCGGUGGCAUGACCACUACUAGUGGCCCAUUGCGG  1534
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
254  GlyTrpLeuAlaAspPheLeuLeuThrIleTrpIleValGlyLeuIleThr  269
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1535  GGCCAGANUGCGGAGAUUUCUUGCGGAGCGCGCCGACNAGUUCACACACCA  1584
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
270  ValArgIleIleAlaThrIleLeuGlySerLeuProSerSerAlaLeu  286
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1585  CGUGCGCAAGUUGAGACACUUGCGGAGCGUUGCGAGUAGAGCACACGUCG  1633
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
286  IeValSerLeuProTrpTrpLeuAsnSerGlyTrpIleThrAlaThrAlaLeu  302
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1635  UGUUGUGUGUGCGGACACCGCACUCCCAAGGGCGUG...GCCAUUCCUUC  1681
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
303  LeuThrLeuSerCysGlyLeuSerThrIleCysGlnSerGlyIleTrpIle  319
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1682  CUGGUCUAGCGCGUGGCGUUCACGCGCUCUUGCCACUUCUUGGUGUACGU  1731
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
319  eAsnValLeuAspIleAlaProArgTrpSerSerPheLeuMetGlyAlas  336
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1732  GAACACACUUGGACAUAGCCCCCGCUGACGCCAGCUCUACUGAGGACUUC  1781
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
336  eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly  352
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1782  CCACGGCGCGUGGGCACACUUGCGGGACUAGUGUUGGCCCAUACUUCGGG  1831
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
353  PheLeuLeuSerGlnAspProGluPheGlyTrpArgAsnValPhePheLe  369
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1832  GCCACUGACUAAACCAACAAGUCUGGGAGAGAGUGGCGACUAGUGUUCUUAU  1881
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
369  uLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTrpTrpLeuIlePheGly  386
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1882  UGCUCUCCUGUGGACUAGUGAGGUGUUAUUCUUCACGGGCGUUCGU  1931
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
386  IuAlaAspValIleGluTrpAlaTrpAlaTrpGluArgGlyLeuThr  399
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||
1932  CUGGACAGGAAGCGCGGUGGGGAGAGCCUUGAGAGAUAGGC  1972
    |||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

```

```

seq_name: /cogn2_6/ptodata/1/1na/PCRTUS_COMB.seq:pct-US96-05792-1
seq_documentation_block:
  Sequence 1, Application PC/TUS9605792
  GENERAL INFORMATION:
    APPLICANT: NI, Binhui
    APPLICANT: Paul, Steven M.
    TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
    TITLE OF INVENTION: PHOSPHATE COTRANSPORTER
    NUMBER OF SEQUENCES: 3
    CORRESPONDENCE ADDRESS:
    ADDRESSEE: Eli Lilly and Company
    STREET: Lilly Corporate Center
    CITY: Indianapolis
    STATE: Indiana
    COUNTRY: United States of America
    ZIP: 46285
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Patentin Release #1.0, Version #1.30
    CURRENT APPLICATION DATA:
    APPLICATION NUMBER: PCT/US96/05792
    FILING DATE:
    CLASSIFICATION:
    PRIOR APPLICATION DATA:
    APPLICATION NUMBER: 08/430,033
    FILING DATE: April 27, 1995
    ATTORNEY/AGENT INFORMATION:
    NAME: Gaylo, Paul J.
    REGISTRATION NUMBER: 36,808
    REFERENCE/DOCKET NUMBER: x-10006

```

```

: TELECOMMUNICATION INFORMATION
: TELEPHONE: (317) 276-0756
: TELEFAX: (317) 276-3861
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 2716 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: CDNA
:
: FEATURE:
:
: NAME/KEY: CDS
: LOCATION: 461..2143
:
: PCT-US96-05792-1

```

alignment_scores:	
Quality:	479.00
Ratio:	2.056
Percent Similarity:	58.690
alignment_block:	
Length:	397
Gaps:	7
Percent Identity:	28.212

Align seg 1/1 to: PCT-US96-05792-1 from: 1 to: 2716

[illegible]

```

204 LeuGLYCysPheSerHisSGINtrPLeuValSerThrMetValValTyrII 220
      ::::: ||| :::::||||
1385 GTGGCCAACTTTCGCCAGCTGGAGCTTCTCACTGCTGCATCTCCCA 1438
      ::::: ||| :::::||||
220 eProrhTyrIleSerSerValTyrHisValAsnIleArgAspAsnGlyL 237
      ||| ||| :::::||||| ::::: ||| |||
1435 GCCCGCACTACTTCGAAGAAGAGTTCGGCTTCGACATCGAAGAGAGGCC 1488
      ::::: ||| :::::||||
237 euLeuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
      ||| :::::||||| ||| ::::: ||| ::::: |||
1485 TGGTGTCCGGGGTCCCGCCACCTGGTGTATGACCAATCATCTGCCATCGCG 1534
      ||| :::::||||| ||| ::::: ||| ::::: |||
254 GLYTyrLeuAlaAspPheLeuLeuThrLysLysPheArgLeuIleThr . 269
      ||| ||| :::::||||| ::::: ||| ::::: |||
1535 GGCCAAATGCGCGACTTCCTCTGGAGAGCCGCCGCATCATGTCCACCACTAA 1588
      ::::: ||| :::::||||| ::::: ||| ::::: |||
270 .ValArgLysIleAlaThrIleLeuGlySerLeuProSerSerAlaLeuI 286
      ||| ||| ::::: ||| ::::: ||| ::::: |||
1585 CGGGCCAAATGTGATGATAGCTGGAGGCTTCGGCATGGAACCAACCGCTGC 1638
      ::::: ||| ::::: ||| ::::: |||
286 IeValSerLeuProTyrLeuAsnSerLysTyrIleThrAlaAlaLeu 302
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1635 TGTGTGTGGTGGCTACTCTCCACATCCCAAGGAGCGTG...GCCATCTCTTC 1681
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
303 LeuThrLeuSerCysGlyLeuSerThrLeuGlySGINserGlyIleTyrII 319
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1682 CTGTGCTCTAGCCCTGGGCTTCACGCGCTTGCATCTCTGGGTTCACAGCT 1731
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
.319 eAsnValLeuAspIleAlaProArgTyrSerSerPheLeuMetGlyAlas 336
      ||| ||| ::::: ||| ::::: ||| ::::: |||
1732 GAACCACTGTGACATAGCCCCCGGCTCGACGACATCTCTATGGGACATCT 1781
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
.336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly 352
      ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1782 CCACAGCGCTGGGCACACTGTCCGGCGTGGTGTCTCCCATCATCTGTGGGG 1831
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
353 PheLeuLeuSerIleAsnProGluPheGlyTyrPargAsnValPhePheLe 369
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
1832 GCCATATCACTAGACACAAGACTCGGAGAGAGTGGCAGTACGTGCTTCAT 1881
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
369 uLeuPheAlaValAsnLeuGlyLeuLeuPheTyrLeuIlePheGlyG 386
      : ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
1882 TGGCTCCCTGGTCCACTATGATGAGGTGTCACTTCCTACGGGGCTTTCCTT 1931
      ::::: ||| ::::: ||| ::::: ||| ::::: |||
386 IAlaAspValGlnGluTrpAlaLysGlnArgLysLeuThr 399
      ::::: ||| ||| ::::: ||| ::::: ||| ::::: |||
1932 CTGGAGAGAAGCAGCCGTGGCGAGAGCTTAGGAGATGAGAC 1972
      ::::: ||| ||| ::::: ||| ::::: ||| ::::: |||
seq_name: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq:PCT-US96-05792-3
seq_documentation_block:
: Sequence 3, Application PC/TUS9605792
GENERAL INFORMATION:
APPLICANT: NI, Binhui
APPLICANT: Paul, Steven M.
TITLE OF INVENTION: HUMAN BRAIN SODIUM DEPENDENT INORGANIC
TITLE OF INVENTION: PHOSPHATE CORRANSROTHER
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05792

```

```

1      FILING DATE:
2
3      CLASSIFICATION:
4
5      PRIOR APPLICATION DATA:
6
7      APPLICATION NUMBER: 08/430,033
8
9      FILING DATE: April 27, 1995
10
11     ATTORNEY/AGENT INFORMATION:
12
13     NAME: Gaylo, Paul J.
14
15     REGISTRATION NUMBER: 36,808
16
17     REFERENCE/DOCKET NUMBER: X-10006
18
19     TELECOMMUNICATION INFORMATION:
20
21     TELEPHONE: (317) 276-0756
22
23     TELEFAX: (317) 276-3861
24
25     INFORMATION FOR SEQ ID NO: 3:
26
27     SEQUENCE CHARACTERISTICS:
28
29     LENGTH: 2716 base pairs
30
31     TYPE: nucleic acid
32
33     STRANDEDNESS: single
34
35     TOPOLOGY: linear
36
37     MOLECULE TYPE: RNA
38
39     CUST-0596-05792-3

```

[illegible]

```
alignment_block:
US-09-391-958-1 x PCT-US96-05792-3 .
```

Align seg 1/1 to: PCT-US96-05792-3 from: 1 to: 2716

```

12 ValProLeuGlySerSerIleAlaArgGlyIleAlaLeuValLeuHisPhe 28
   :::::::::: ::::::::::::::::::::
941 AUCGCCCAACGUCGCGCGGCCACAGUAGCGUCUGGACAC..... 979
   ::::::::::::::::::::

28 eCysAsnPheThrThrIleAlaGlnAsnValIleMetAsnIleThrMetV 45
   ::::: ||| |||::: ::::: ::::: |||
980 .....UUCGAGAGAUUCUCCAGGAGGUGUGAGAGGAGGACACAUACC 1022
   ::::::::::::::::::::

45 AlaIleMetValAsnSerThrSerProGlnSerGlnLeuAsnAspSerSer 61
   |||
1023 CCGCCCGCC..... 1030
   ::::::::::::::::::::

62 GluValIleuProValAspSerPheGlyGlyLeuSerTyrAlaProLysE 78
1030 ..... 1030
   ::::::::::::::::::::

78 IleuProAlaLysSerSerIleLeuGlyGlyGlnPheAlaIleTrpGlnA 95
1031 .....CAUGGAGUUCGAGACA 1046
   ::::::::::::::::::::

95 rGfrPrpLysProProGlnGluArgSerArgLeuCysSerIleAlaLeuSer 111
   ::::::::::||| ||| ||| ||| ||| ||| ::::: |||:::
1047 AAGGGCCCCCACCUCUAGACGGAAGUCCCGUCGACACAGACCCUUCUG 1096
   ::::::::::::::::::::

112 GlyMetLeuLeuGlyGlyCysPheThrAlaIleLeuIleGlyGlyPheIleS 128
   ||| |||::: ::::: ::::: ::::: |||:::
1097 GGUCCUACUGCGGGGCGGUGUCCGACAGCCCCCGCGGGGUCUUCUG 1146
   ::::::::::::::::::::

128 rGluThrIleuGlyTrpProPheValPheTyrIlePheGlyGlyValGlyC 145
   ::| ||||| ||||| ||||| ||||| ::::: |||
1147 GCGAGUACGAGAUAGAGUCUUGUUCUACGUCUACGCGACGACUUCGGGA 1196
   ::::::::::::::::::::

145 ySValCysSylLeuLeuThrPheValValIleTyrAspAspProValAsr 161
   |||::: |||::: |||::: |||::: |||::: |||:::
1197 UCUUCUGUACUCUGUUCGCGUCGUCGUCUCCUACGAGUCCCGCGCGUG 1246
   ::::::::::::::::::::

162 TyrProTrpLysSerThrSerGluTyrGlySerTyrIleIleSerSerLeu 178
   ::::: ||||| ||||| ::::: ||||| ||||| :::::
1247 CACCCAGAGAUUCGAGAGAGAGGACAGACAGUACAC.....GA 1294
   ::::::::::::::::::::

178 sGlnGlnValGlySerSerTyrGln.....ProLeu..... 188
   ::::::::::::::::::::

```

```

1285 GAGCGCAUCCGAGAGAGCGCAACUACUAGAACCCCGUACGAGUUUA 1334
189 ....ProilelysaAlaMetleuArSerleuProiletpSerileCys 203
1335 GCACUCCCGCGCGCGCUUCACCGCUUAGUGGAGUCUAGCCAUAC 1384
204 leuGlyCysPheSerHisGlnTrpLeuValSerThrmethValValTyr 220
1385 GUGGCGCAACUCCGCGCACGUCGAGGUCUCCUGGUCUACUCCCA 1434
220 eProthTyrIleSerSerValTyrHisValAsnIleArgAspAsnGly 237
1435 GCCCGCAUACUUCGAGAGAGUUCGCGCUUCGAGAUACGAGAGUAGCC 1484
237 euleuSerAlaLeuProPheIleValAlaTrpValIleGlyMetValGly 253
1485 UGUUGUCCGCGCGCGCACCUUGUCUAGCAUACUGGCCAUCCGCGC 1534
254 GlyTyrIleuAlaAspPheLeuLeuThrLysLysPheArgLeuIleThr 269
1535 GGCAGAUCCGCGAGACUCCGCGAGCGCGCAUCCAUUGCCACCA 1584
270 .ValArgLysIleAlaThrIleLeuGlySerleuProSerSerAlaLeu 286
1585 CGUGCGCAUUGUAGUACUGCGGAGGCUUCGCGCAUAGCAACCGCUGC 1634
286 leValSerleuProTyrLeuAsnSerGlyTyrIleThrAlaThrAlaLeu 302
1635 UGUUGGUGGCGCGCUUCCGCAUCCCAAGCGCUG...GCCAUCCUCC 1681
303 LeuThrLeuSerCysGlyLeuSerThrLeuGlyGlnSerIleTyrIle 319
1682 CUGGUCUCCAGCCGUGGCUUACGCGCUUCGCCAUCCUGGUCACAGU 1731
319 easnValIleuAspIleAlaProArgTyrSerSerPheLeuMetGlyAs 336
1732 GAACCAACUUGGACAUAGCCCGCGCUACGCGCAGAUCCUACUGGCAU 1781
336 eArgGlyPheSerSerIleAlaProValIleValProThrValSerGly 352
1782 CCAACGCGCGGCGCACUUCGCGGCUUGGUGGCCAUCCAUCCUGGCG 1831
353 PheLeuSerGlnAspProGlnPheGlyTrpArgAsnValPhePheLe 369
1832 GCCAUGGCUAGCAACAAGACUCCGAGAGAGUGGCAUGAGUCCUAU 1881
369 uLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuIlePheGly 386
1882 UGCUCUCCUGGUCACUAGGAGUGUUCACUCCUACGCGGCUUCUUCU 1931
386 luAlaAspValGlnGluTrpAlaLysGluArgLysLeuThr 399
1932 CUGGAGAGAGAGCGCGGCGCAGAGCCUUGAGGAGAGAGUAGAC 1972
seq_name: /cgn2_6/ptodata/1/lna/5D_COMB.seq:US-08-805-118-5
seq_documentation_block:
; Sequence 5, Application US/08805118
; Patent No. 5985604
; GENERAL INFORMATION:
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: NOVEL HUMAN SODIUM-DEPENDENT
; TITLE OF INVENTION: PHOSPHATE CO-TRANSPORTER
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/805,118
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0221 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITV02
; CLONE: 754412
US-08-805-118-5

alignment_scores:
Quality: 433.00 Length: 85
Ratio: 5.094 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-391-958-1 x US-08-805-118-5 ..

Align seq 1/1 to: US-08-805-118-5 from: 1 to: 272
317 11eTyrIleAsnValIleAspIleAlaProArgTyrSerSerPheLeuMe 333
|||||
1 ATTTATATCAATGCTTATGATATTGCTCCAAAGTATTCACATTTCAT 50
333 tGlyAlaSerArgGlyPheSerSerIleAlaProValIleValProThrV 350
|||||
51 GGGAGCATCAAGAGATTTTCGACATAGACACCTGTCATTGTACCACTG 100
350 a1SerGlyPheLeuLeuSerGlnAspProGlnPheGlyTrpArgAsnVal 366
|||||
101 TCAATGATATTCCTTATGTCAGACCCCTGAGTTGGGTGAGGAAATGTC 150
367 PhePheLeuLeuPheAlaValAsnLeuLeuGlyLeuLeuPheTyrLeuI 383
|||||
151 TTCTTCTTGCTGTTTGCCTTAACCTGTTAGGACTCTTCTACCTCAT 200
383 ePheGlyGluAlaAspValGlnGluTrpAlaLysGluArgLysLeuThra 400
|||||
201 ATTTGAGAGAGACAGATGTCACAGAAATGGCTAAAGAGAGAAATCACTC 250
400 rGleu 401
|||||
251 GTTAA 255

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-724-394A-20
seq_documentation_block:
; Sequence 20, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Krommal, Gregory S.

```

APPLICANT: Lauer, Peter M.
APPLICANT: Ruddy, David A.
APPLICANT: Thomas, Winston
APPLICANT: Tsuchihashi, Zenta
APPLICANT: Wolff, Roger K.
TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
TITLE OF INVENTION: Sequences and Antibodies Thereeto
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSER: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 20:
SEQUENCE CHARACTERISTICS:
LENGTH: 246240 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..246240
OTHER INFORMATION: /note= "HLA-H. CONTIG"
US-08-724-394A-20
alignment_scores:
Quality: 297.50 Length: 543
Ratio: 1.848 Gaps: 10
Percent Similarity: 29.650 Percent Identity: 20.994
alignment_block:
US-09-391-958-1 x US-08-724-394A-20/rev ..
Align seg 1/1 to reverse of: US-08-724-394A-20 from: 1 to: 246240
137 PheYrIlePhe.....GlyGlyValGlyCysValGlyCysLeuLeuTr 151
|||||
22330 TTTTATCCCTTTCTCCAGAGTAGCAGCTGCTGCTGCTGCTGCTGCT 22281
151 pHeValValIleTyAspAspProValSerTyProTrPLeSerThrs 168
|||||
22280 GTTCACAGGATTATGATGAGCCCATGCATGACCCGTCATTAAGTGTTA 22231
168 eGluGlyGluTyTrIleIleSerSerLeuLysGlnGlnValGly..... 182
|||||
22230 GGGAAAGAGACATCTGCTCCTCAGTGGCACAACAGGTACAGTGACA 22181
183Se 183
22180 CCTTACTCTGGCCCATGACAGAGCTCTAGAGGAGGTGTGATCTC 22131
183 rSerLysGln..Pro..... 187
|||||

22130 CTCTGAGAGGACACACATCTTGGTGTCTATTAATCAATGCTGATTAGATCT 22081
188LeuProIleLysAlaMetL 194
22080 TCTTTTTCACCCAGTCTCTCCAGAGAGCTGTGCCATTAAGCCATG 22031
194 euaIrsSerLeuProIleTrpSerIleCysLeuGlyCysPheSerHISgln 210
:::|||||
22030 TCACATGCCCTACACACTTTGGCCATTTTCTGGGTTTTCACGCCATTC 21981
211 TrpLeuValSerThrMetValValTyTrIleProTrhTyTrIleSerSerVa 227
|||||
21980 TGGTGTGACACATCATCTTAACATACATACATACATACATACATACAT 21931
227 LTyHISValAsnIleArgAsp..... 234
:::|||||
21930 GCTCCATGTTAATCATCATGAGATGTGATTACTTCTATACCTTCAGAA 21881
234 234
21880 AATGATATATGTTAATAGGAGAAAGAGTTCGTGTACTTACTTACTTCT 21831
234 234
21830 GGCTTACATATTAACCATTAATTTAATCTTCAATAGCTTGAGAGAG 21781
234 234
21780 CATGTTTATTAATTCCTTTTTCACAGATGTGAAACAGCACTTACAGGT 21731
234 234
21730 GAGATTAATTTGCCCGAGTTTGACAAATTAAGTATAGAGCTGTGCAG 21681
234 234
21680 CATCATATTTCTTAACCACTATGCTATTAATCAACACAGCTGATTCNA 21631
234 234
21630 AGCTTCTTTAATAAATAATATTGCTGGCCAGGACATGTGTATGCTG 21581
234 234
21580 TAATTCACACACTTTGGAGGCCGAGGACGACATCATAGAGTACAGAA 21531
234 234
21530 TGCAAGACCAAGCTGACCAATATGTTTACTAATATCATCTTAATAAA 21481
234 234
21480 TACAAAAATTACCAAGGTGTGTGTGACAGGACACTGTAATCCAGCTATTC 21431
234 234
21430 AGAGAGCTGAGACAGAGAGAAATGCTTGAACCCAGAGAGGTGAGTTGAT 21381
234 234
21380 TGAGCAAGATCATCTCACTGCATCTCAAGCTGGGCGACAGATGAAGCT 21331
234 234
21330 CCGTTTCAAAAAAACAACCAAGAAATTAATATTGCTTTATCTGAGAG 21281
234 234
21280 CCAAGATGATGACAGCTTGTGGCCCTTTATCTGAGACAGTGTCTTTTA 21231
234 234
21230 GTGTGAAAAAGAGATGCTAATTTTCCCAACCAACCAAGTATCATGAG 21181

```

234 ..... 234
21180 GGTAAGTTAAATGCGTCGTCTGTACTGACAAATTTGGTCTTAACGTA 21131
235 ..... Asn..GlyLeuLeuS 239
21130 TCTGTATACTACTGTGTATAAATCTTCCTCTCAGAGTGGAGTTCTGT 21081
239 eFALeupProPheIleValAlaTTPValILleGlyMetValJgLygLyTyr 255
|||||.....
21080 CCTCCCTGCTTTTATTTGCTGCTGCAAGCTGTACAAATTTTAGAGGTGAG 21031
256 IeuAlaAspPheIleLeuThrIlysLys...PheArgIleIleThValAr 271
|||||.....
21030 CTGGAGATTTCTCTTTTGTCCAGGAATCTTCTCAGATGTATCATCTGTGG 20981
271 GlySILeAlaThrILleLeuGlySer..Leu..... 280
20980 AAAGCTCTTTTCATCTCTCTGTGAAGATTAAGGCTGTGGCCCATTTTANC 20931
280 ..... 280
20930 AATCCCTTTTCTGCACATGCTCAGAGGTTCCCTGACAGCATGCTCTC 20881
281 ..... ProSer..SerAlaLeuILleValSerLeuProTyr 291
20880 AATGCCCAAGGCTCCTCCTTCATCATAAATATGTGCTGTGGCCTGCCCCTT 20831
292 IeuAsnSerGlyTyrILleThrAlaThrAlaLeuLeuThrLeuSerCysG 308
::: |||::: ||::: ||| ||| |||
20830 GTGGCTTCAGATTGATGATTAACCAATTAATTTTGTGATATCACTTAATCCNG 20781
308 yLeuSerThrLeuCysGInSerGlyILleTyrILleAsnValILleuAspILe 325
| |||::: |||::: |||::: |||::: |||::: |||::: |||::: |||
20780 GACCACTAACCTATGTGACTCAGGGTTTATTCATCAACACCTTAGATATCG 20731
325 IAPoArgTyrSerSerPheLeu 332
|||||::: |||::: |||
20730 CCCCCAGGTAAAGCTCTACCTG 20708

seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-724-394A-21
seq_documentation_block:
: Sequence 21, Application US/08724394A
: Patent No. 5872237
: GENERAL INFORMATION:
: APPLICANT: Feder, John N.
: APPLICANT: Krommal, Gregory S.
: APPLICANT: Laufer, Peter M.
: APPLICANT: Ruddy, David A.
: APPLICANT: Thomas, Winston
: APPLICANT: Tsuchinashi, Zenta
: APPLICANT: Wolff, Roger K.
: TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e)
: TITLE OF INVENTION: Sequences and Antibodies Thereo
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
: STREET: Two Embarcadero Center, 8th Floor
: CITY: San Francisco
: STATE: CA
: COUNTRY: USA
: ZIP: 94111-3834
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/724,394A
: FILING DATE: 01-OCT-1996
: CLASSIFICATION: 536

```

```

? ATTORNEY/AGENT INFORMATION:
?   NAME: Fitts, Renee A.
?   REGISTRATION NUMBER: 35,136
?   REFERENCE/DOCKET NUMBER: 017957-000100
? TELECOMMUNICATION INFORMATION:
?   TELEPHONE: 415-576-0200
?   TELEFAX: 415-576-0300
? INFORMATION FOR SEQ ID NO: 21:
? SEQUENCE CHARACTERISTICS:
?     LENGTH: 246240 base pairs
?     TYPE: nucleic acid
?     STRANDEDNESS: not relevant
?     TOPOLOGY: not relevant
? MOLECULE TYPE: cDNA
? FEATURE:
?   NAME/KEY: misc.feature
?   LOCATION: 1..246240
? OTHER INFORMATION: /note= "HLA-H.COMFIG
US-08-724-394A.21

```

```

alignment_scores:
  Quality: 297.50      Length: 543
  Ratio: 1.848         Gaps: 10
  Percent Similarity: 29.650      Percent Identity: 20.994

alignment_block:
  US-09-391-958-1 x US-08-724-394A-21/rev ..

Align seg 1/1 to reverse of: US-08-724-394A-21 from: 1 to: 2462240

```

137	PheTyrIlePhe..... GlyIValGlyCysValCysLeuLeuTr	151
22330	TTTATTCCTTCTTCCCAAGTAGACACTGGCTGTGCTGCTCTCTTAAG	22287
151	pPheValValIleTyrAspAspProValSerTyrProTrpIleSerThrS	168
22280	GTTTCACAGAGATTATATGATGACCCCATGCATGCCCGCATAAAGTGTTA	22233
168	ergLIuLlgluTyrIlelleSerSerleuLIyngInValGIy.....	182
22230	GGAAGAAGAGACATCCGTCTCTCACTGGCTCAACAGGTACAGTGCACA	22183
183Se	183
22180	CCTTGACTCTGGCCCATGACAGAGTCTCTAGGCGAGGTGTGATCTC	22133
183	rSerLygLn..Pro.....	187
22130	CTCTGAGAGGACCAATCTGGCTGCTAAATCTCATGCTGATTAATCT	22083
188LeuProIleLysAlaMetL	194
22080	TTCCTTTGAGCCCAAGTCTCTCGACGAGCTGTCCCCATAAAGGCGATGG	22033
194	euaGrSerleuProIleTrpSerIleCysLeuGIYCysPheSerHisLn	210
22030	TCACATGCGCTACACTTTCGGGCAITTCCTGGGTTTTTTCAGCCATTTTC	21983
211	TrpleuValSerThmEtaValTyrIleProTrhTyrIleSerSerVa	227
21980	TGGTTGTGCACCATCATCTCTTAACATACACCACAGATATATCAGTACTCT	21933
227	LTYrHisValasnIleArgasp.....	234
21930	GCTCCATGTTACATCAGAGATGTAGTTTACTTCTCTAATCTTACGAA	21883
234	234
21880	AATGATAATGTAATAAGAGAAAACAGTTCTGTGTACTATTACATTCCT	21833
234	234

21830 GCGTTACATATATACATTAATTTACCTTCACATGACCTTGAGAGG 21781
 234
 21780 CATGTTATATATTCCTTTCACAGATGTGGAACAGAGACTTAGAGT 21731
 234
 21730 GAGATACTTGCCCGAGTTGACAAATACTAAGTATAGAGCTGCTCAG 21681
 234
 21680 CATCCATATTTCTTAACCATATGCTATTAACACACAGCTGATTCGA 21631
 234
 21630 AGCTTTCTTTAGAAATATATGCTGGGCCAGGACATGCTGTCCTG 21581
 234
 21580 TAATTCAGCACTTTGGAGCCGAGCAGACATCATGAGTCAAGAA 21531
 234
 21530 TGCAGACACAGCTGACCAATATGTTTACTAATATCATCTACTAAAA 21481
 234
 21480 TACAAAAATTAGCCAGGTGTGTGGCAGGACCTGTATCCAGCTATTC 21431
 234
 21430 AGGAGCTGAGACAGAGAAATCGTTGAACCCAGAGGTGAGTTGCAT 21381
 234
 21380 TGAGCCAAATCATGCTGCACTGCACTCAGCCTGGCGACAGAGTAAGT 21331
 234
 21330 CCGTTTCAAAAACAAAAAACCAAGAAATTAATTTGCTTTATCTGGAG 21281
 234
 21280 CCCAGAGTATGACGCTTGTGCTTATCTGAGACAGTGTCTTTTA 21231
 234
 21230 GTGTGAAAAAGATGCTAATTTTCCCCCAACACCAAGATATCGGG 21181
 234
 21180 GGTAAATTAATGCTGTGTGTACTGACAAATTTGGTGTAACTGA 21131
 235
 21130 TCTCTATTAACCTCTGATAACTCTCTCTCCTCAGAGTGTGATCTGT 21081
 239
 239 erAlaLeuProPheIleValAlaTrpValIleGlyMetValGlyGlyTyr 255
 ||:|||||
 21080 CCTCCCTGCTTTTATGCTGTGCAAGTGTACATTTTAAAGAGGTGAG 21031
 256 LeuAlaAspPheLeuLeuThrIleLys...PheArgLeuIleThrValAla 271
 |||||
 21030 CTGGCAGATTTCTTTTGTCCAGGAATCTTTCAGATGATCAGCTGGG 20981
 271 GlyIleLeuThrIleLeuGlySer.Leu..... 280
 |||||
 20980 AAAGCTTTTTCATCTCTGTGAAGATAAGCGTGTGGCCCATTTAAC 20931
 280
 20930 AATCCCTTTTTCGACATGCTGTCAGAGGTTCCCTGACAGATGTCTC 20881

281ProSer.SerAlaLeuIleValSerLeuProTyr 291
 |||||
 20880 ATTGCCAGGCTCTCTCTTCATCATATATGCTGTGCGCCCTT 20831
 292 LeuAsnSerGlyTyrIleThrAlaThrAlaLeuThrLeuSerGly 308
 ||:|||||
 20830 GTGGCTCCAGTATGATGATACCATATTTTCTGATCTATTCCTGG 20781
 308 yLeuSerThrLeuGlySerGlyIleTyrIleAsnValLeuAspIle 325
 |||||
 20780 GACCAGTAACTTATGTGATCAGGTTTATCATCAACACTTGAATATG 20731
 325 LAbProArgTyrSerSerPheLeu 332
 |||||
 20730 CCCCAGGTAAAGCTTACTCTG 20708
 seq_name: /cgn2_6/ptodata/1/lna/5C_COMB.seq:US-08-724-394A-22
 seq_documentation_block:
 ; Sequence 22, Application US/08724394A
 ; Patent No. 5872237
 ; GENERAL INFORMATION:
 ; APPLICANT: Feder, John N.
 ; APPLICANT: Krommal, Gregory S.
 ; APPLICANT: Lauer, Peter M.
 ; APPLICANT: Ruddy, David A.
 ; APPLICANT: Thomas, Winston
 ; APPLICANT: Tsuchihashi, Zenta
 ; APPLICANT: Wolff, Roger K.
 ; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237e1
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
 ; STREET: Two Embarcadero Center, 8th Floor
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94111-3834
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/724,394A
 ; FILING DATE: 01-OCT-1996
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Pitts, Renee A.
 ; REGISTRATION NUMBER: 35,136
 ; REFERENCE/DOCKET NUMBER: 017957-000100
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-576-0200
 ; TELEFAX: 415-576-0300
 ; INFORMATION FOR SEQ ID NO: 22:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 246240 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: not relevant
 ; TOPOLOGY: not relevant
 ; MOLECULE TYPE: cDNA
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: 1..246240
 ; OTHER INFORMATION: /note= "HLA-H.COMTIG"
 ; US-08-724-394A-22
 alignment_scores:
 Quality: 297.50 Length: 543
 Ratio: 1.848 Gaps: 10

21530 TCGAAGACCAGCCTGACCAATATGGTTTACTAAATATCATCTACTAAAAA 21481

234 234

.....

.....

.....

QUESTION

331

0123456789ABCDEFGHIJKLMN O PQRSTUVWXYZ [\] ^ _ ` { | } ~ ! " # \$ % & ' () * + , - . / : ;

332

[illegible]

C

[illegible][illegible][illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

[illegible]

2
2
4
1
2
2
3
1
4
2
2
2
2

[illegible]

333

[illegible][illegible][illegible][illegible][illegible]

308 Wt auGcmphat auGmnc} nScawc]wt] oTtwrt] aAavva}r awAvart] ah 396:

[illegible]

325 Introduction to Corrosion 333

00730 ||||| :: :||| 007A9
00730 ||||| :: :||| 007A9

D	b	121	IFIPQAAVGEWTIVICRVVQYIGTIOGVITVAHQHEIHWKNAKPLEBERGLTSMLSGFLDLP	180
Q	y	118	FTALLIGGFISELGMFVFYVITIRGGCVCCCLLMFVYIDDDPVSYWIMTSISEKEYITSSL	177
D	b	181	FYTLVLGIIICESGAMWVVFYIFGACGCAACLLMLFVYLDYDDPDHCVSLSHEKEYITSSL	240
Q	y	178	KQVGGSKOPLPIKAMLRSLPIINSICLGFCSHOMLVSTWVVFYPTFYISSYVYHNIRDGL	237
D	b	241	IQGSSSTROSLPIKAMKLSPLNIAISPCFAYLMTVYSRLYVPTLINSMLHDYIRENGI	300
Q	y	238	LSALPEIVAWVIGWGVXIADPELLTKFKRLIT-VRKIATILGSLPSSALVSLPYLNSGY	296
D	b	301	LSSLPYFPAWICGVYIAHVTADPLMSRNMUSLTVIRKLEITAIIGLLPIVSMCLLYISSGF	360
Q	y	297	ITATALLTISCGSLTIOSGYIVNVLDIARVYSSFTMGASRGFSIAPVIVPTVSGFLDS	356
D	b	361	YSTTFPLINASSSFCIGALINADLAPRYVVFYIKGVTTILGTMGKMTSVAGLFTS	420
Q	y	357	ODPEGRANVFELFANVLIGLFPYLIFEGADVOENAKPEKRLRL	401
D	b	421	QDPSSSMFKIFLLMSITIVSVYTFLEFAAEIODMAKKEQOHRRL	465

RESULT 5
I59302
brain specific Na⁺-dependent inorganic phosphate cotransporter - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence-revision 02
C:Accession: I59302
C:NL, B.: Rosteck, P. R.; Nadi, N. S.; Paul, S. M.
Proc. Natl. Acad. Sci. U.S.A. 91, 5607-5611, 1994
A:Title: Cloning and expression of a cDNA encoding a brain-specific Na(+) -dependent inor
A:Reference number: I59302; MUID:94261635
A:Accession: I59302
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-560 <RES>
A:Cross-References: EMBL:U07609; MID:9507414; PIDN:AA19646.1; PID:9507415

Query Match	22.9%;	Score 475;	DB 2;	Length 560;
Best Local Similarity	28.0%;	Pred. NO. 2.2e-30;		
Matches 111;	Conservative 66;	Mismatches 158;	Indels 62;	Gaps 6

```

OY      12 VPSLCSARGLALYVHFCNFTTIAONVIMINIMVAMVNSTSPQSOLDNSEVLPVDSFGG 71
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      161 IPSAARVHGCYI-----FVAILQGLVEGVTYPPAC-----191

OY      72 LSKAPKSLPAKSSILGOGFAIWMERMGPPQERSRLCSIALSGMLGCFWAILIGGFISETL 131
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      191 -----HGIMSKNAPPLERSRLAFTTACGSAAGAVVAMPPLAGVLYQVS 232

OY      132 GMPFVFYIFGCVGVCCLLMFVYIYDDPVSYWIMSTSEKEYIITISLAKQOVGSS----- 185
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      233 GWSVFYFYGGSGIFWYLFWILVSYESPALHPHSISEEEKRYI----EDAIGESAKMLNPV 288

OY      185 -KQPIKAMLRSLPIMSTCLOCFSHQMLVSTWVYIPIYISSYVHVNIIRDNGLSALPLP 243
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      289 TRKENTPMRRFTSMVYVYAIYANFCRSMFTYLLISOPAYFEFVEGFEETSKVGLSALPH 348

OY      244 IVAWYIGWGVGLADPLLTTRKFRLLT-VKIKIATILGSLSPSSALIVSLPPLNSGYIATAL 302
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      349 LVMITIVYIGGDIADPLNSRHMSTNNRKLMNCGGFGHEATLLLVYGSISHKGV-AI5FP 407

OY      303 LTLSCGLSTLCOGIYINVLDIAPRYSFLLMGASRGFSSIAPIVPTVYSGFLLSODPEFG 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      408 LVLAWFGSGFALISGENVNHLDIAPRYASTILMCSINCVGLTSGWCPILVIGAMTKHKTRTE 467

OY      363 WRNPFLLFAVNLGLLEFYLIGEDADVQEMAEKRILT 399
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      468 WQYVFLIASLVHVGVIIFGVAFSGEKKPMAPEEHS 504

```

RESULT 6
T23589
hypothetical protein K10G9.1 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 29-Oct-1999
C:Accession: T23589; T24636
R:Morlmore, B.
submitted to the EMBL Data Library, August 1994
A:Reference number: Z19765
A:Accession: T23589
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-573 <WIL>
A:Cross-references: EMBL:Z26282; PIDN:CA85289.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone K10G9
R:Buck, D.
submitted to the EMBL Data Library, February 1995
A:Reference number: Z19915
A:Accession: T24636
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-573 <WT2>
A:Cross-references: EMBL:Z48055; PIDN:CA88135.1; GSPDB:GN00021; CESP:K10G9.1
A:Experimental source: clone T07A5
C:Genetics:
A:Gene: CESP:K10G9.1
A:Map position: 3
A:Introns: 38/3; 87/3; 224/1; 272/3; 322/3; 444/3; 476/3; 496/1; 517/2

Query Match	21.6%	Score 449	DB 2	Length 573
Best Local Similarity	26.7%	Pred. No. 2.6e-28		
Matches 125	Conservative 84	Mismatches 165	Indels 94	Gaps 14

```

Oy      1  MOVDETLIPRKVPSSCSAYGIALYLF-----CNETTIAQVINNTITVAANVST  51
Db      35  LQEDKKFLRKY-----RMQIALLAHREFALSPGRSNF-GVAKRNM-----NFT  80
Oy      52  SPQSLND-----SSEVLVDS--FGLSK-----APKSLPAKSSILGGQFA---  92
Db      81  DAYGEVHEKEFEPMGTETVGMMESEFPYGAAQIPAGVIAAKFAPKRLPMLGILFASLTN  140
Oy      92  -----IHERGQPERKRLCSIALSGML  114
Db     141  IYTAICLNFHPPTDIFVAVIQVMQIALGVCPAMHGVMKVAAPLERSKLATTTFTGAS  200
Oy     115  LGCFPAALIGGFIENLGMPEVFYIEFGVGCYCCLLAFVITVDPPSYMISSEKEYII  174
Db     201  VGVMVGLPASATLVSHFSKSTPFYFGALGYMSILMIYVSGTSPETHYISADEKKYI-  260
Oy     175  SSLKQVGS-----SKQPLPIKAMLRSLPIWSICLGCSHOMLVSTVAVYIPITYSSVY  228
Db     260  ---TEKVGSVAVKNNMTLTLLPWRDMWTSVAWVAIIICSCRCRSMSPFLLLGNQLTWKMDL  316
Oy     229  HVINIDNGLLSLPRTIVAAVIGMVGCIYLADELITLK-KPFLITVKRIATILGSLPSALYI  287
Db     317  HDIDNNGSIALFEPOLGKMIYVTLTSGQLSDYLRSSGKMSTEARVRSVNFEGTVEAVMIG  376
Oy     288  SLPIYNSGITATFALLTSLGSLPQCSGIYINVLDIARVSSPLTMGASRGSSSIAPVY  347
Db     377  CLAFVRDPIAVITFILI-LACSGGAVLSCFNVNHHDIARHAPILMGLINGALAIAGV-G  434
Oy     348  PTVSGFLLSODPEFGMRNVFLLFAVNNLGLLFYLFEGADVQEWAKE  395
Db     435  GIVTNSLGYQNPD-GMQWVFLLMASIDIRGILFPIIFAKGDVILPMARE  481

RESULT
7
T24653
hypochetrical protein T07A5.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24653

```

Query Match	20.98	Score	434	DB	2	Length	544
Best Local Similarity	25.48	Pred. No.	3.9e-27				
Matches	119	Conservative	85	Mismatches	167	Indels	98
						Gaps	13

RESULT 8
S28286
hypothetical protein C38C10.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 12-Mar-1993 #sequence_revision 12-Mar-1993 #text_change 16-Feb-1997
C:Accession: S28286
R:Thomas, K.
Submitted to the EMBL Data Library, December 1992
A:Reference number: S28285
A:Accession: S28286
A:Molecule type: DNA
A:Residues: 1-472 <THO>
A:Cross-references: EMBL:Z19153
C:Genetics:
A:Introns: 50/3, 287/3, 351/3, 412/3
C:Keywords: Transmembrane protein

Query Match	20.78;	Score 429;	DB 2;	Length 472;
Best Local Similarity	27.98;	Pred. No. 8.3e-27;		
Matches 109;	Conservative 83;	Mismatches 162;	Indels 36;	Gaps 12;

RESULT 9
T01534
hypotheical protein A.IG005110.nm - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01534
R:Andrews, S.
submitted to the EMBL Data Library, July 1997
A:Description: The sequence of A. thaliana IG005110.
A:Reference number: Z14347
A:Accession: T01534
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-413 <AND>
A:Cross-references: EMBL:AF013293; NID:g2252823; PID:g2252847
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 79/3; 115/3; 128/3; 200/3; 331/3; 353/3; 382/1
A:Note: A.IG005110.nm

	Query Match	18.7%;	Score 388;	DB 2;	Length 413;	
	Best Local Similarity	25.7%;	Pred. NO. 1.3e-23;			
	Matches	106;	Conservative	69;	Mismatches 135;	Indels 104; Gaps 11.
QY	19 RYGALVHFCNFTTIAONVMITMVAWVNSTPQSO-LNDSESLPV-----	67				
Db	60 RRWIVLLCFESF-----LTCNMRVNNSIALLPMOGEYNSSAKTVGLIQSFFMGYLL	113				
QY	67 -----DSFGLSKAPRSLPAKSSILGGFALWMENGPQERSRLCSIALSGMLGC	117				
Db	114 TQLIGIWADNFGGGVA--MPANNNL-----SKMIPSEKSRIATLYSCMYLGS	163				
QY	118 FTALLIGFISETLGMPEVEYIFGGVGCCVLMEFVIYDDPSVPMISTSEKEKYLISL	177				
Db	164 VTGLAFSPMLTTKGWPSPVFYSFGSLGIWEFLWLTKFAFPYSSRPDPLDEEERKVILGGS	223				
QY	178 KOQGSSKOP---PIRKMLRSLPMTSSICLGFSHOWLVSMNVYIIPTV-----I	224				
Db	224 K-----PREPTVAVPKAILSKPPWALLISHFCINMGSTFILLTMPPYYNOARSSASVT	278				
QY	225 SSVVHV-----NIRDNGLLSALPITVAWVIGAVGYLADFLLTKFRRLTYVKIKITI	276				

```

Db      279 SLTFNIFCEQVLKFWLNTESGLLCPWILMAVFANIGMIADTLVSRG----- 327
Oy      277 LGSLSNALIYSLPLNNGYITATATLTLSCGLSTLCSGIIYINVLADIPRYSFLMGAS 336
Db      327 -----LSTNGSDAFSOSGLXSNQDIDIPRAGVLLGLIS 360
Oy      337 RGFSSIAPIVPTVSGFLLSDPFGMRVFFLLFAVNLGLLFYLFIEAD 388
Db      361 NTAGVLACVFGTAATGCTIIGRG---SMDYFVNAVALYIGLTVNMLPRTGE 409

```

RESULT 10
 724729
 hypothetical protein T09B9.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T24729
 R:Kershaw, J.
 submitted to the EMBL Data Library, December 1994
 A:Reference number: 219929
 A:Accession: T24729
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-516 <MIL>
 A:Cross-references: EMBL:247070; PIDN:CAAB7340.1; GSPDB:GN00028; CESP:T09B9.2
 A:Experimental source: clone T09B9
 C:Genetics:
 A:Gene: CESP:T09B9.2
 A:Map position: 10
 A:introns: 85/1; 143/2; 184/1; 241/3; 344/1; 391/1; 436/2; 470/3

Query Match 15.4%; Score 320; DB 2; Length 516;
 Best Local Similarity 25.4%; Pred. No. 4.5e-18;
 Matches 105; Conservative 78; Mismatches 177; Indels 54; Gaps 13;

```

Oy      29 CN-----FTTAQNYI---MNTIVAMVNSTSPQSOLD-----SSFELPV 66
Db      71 CNHPRQPMWSSIQQGLITISGNGFSLPMVITGMQADRLKMTIYAAMAFTIVSNAVLPT 130
Oy      67 D-----SFG-----GLSKAPKSLPAKSSILGQFAIWRNGPPOERSRLCSIALSGML 114
Db      131 SAGASFALVFLRYLVTGFDALLS-PASSSL-----TRMFPFRERSALGIYTSGRQ 182
Oy      115 LCCFPAALLIGFT-----SETL-GMPFVYITFGVGCVCCLLMFYVITYDDPVSYPISTIS 168
Db      183 IGTLLIPLTGGWLCGSDGSKFLGMPAIFYLSVVAALVIMVVSADKPSKHLICISHN 242
Oy      169 EKEYIISLKOO-VG--SSKQPLPIKAMRLSPIMISICLGCPSHQLVSTMTVVIPTIYS 225
Db      243 EBAVYNRKRIEENIGKRRNRKNTPKKALFTSKOVVAVALVCHPEPLVIMQLPKFPS 302
Oy      226 SVYHAVNIRNDGLSALPFIYAVVIGVGGYLAFLTKKF-RLITVRKIATTLIGSLPSSA 284
Db      303 DVLGSLNFTVNGVLSALPMAILFLSKCLSLASLASYLANGLYLRKTSCKIFNFIASGLGI 362
Oy      285 LIVSLPYLNS--GYTATATLTLSCGLSTLCSGIIYINVLADIPRYSFLMGASRGFSSI 342
Db      363 CIAAAPPLMSNLOHAIWAILIICLANAFGLHRTGVTATVQALAPAFSGIITGIAFAVASC 422
Oy      343 APVIPTVSGFLLSDPFGMRVFFLLFAVNLGLLFIEADVQMAKER 396
Db      423 PSIFKKLISQILRTGSKHEWTIVEISAFVALPLPIFTLWGSARTEMASNR 476

```

RESULT 11
 715201
 hypothetical protein F12B6.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T15201
 R:Pauley, A.; Magg1, L.

submitted to the EMBL Data Library, May 1997
 A:Description: The sequence of C. elegans cosmid F12B6.
 A:Reference number: 218307

A:Accession: T15201
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-499 <PAU>
 A:Cross-references: EMBL:AF003138; NID:92088708; PID:92088710; PIDN:AAB54154.1; GSPDB
 A:Experimental source: strain Bristol N2; clone F12B6
 C:Genetics:
 A:Gene: CESP:F12B6.2
 A:Map position: 1
 A:introns: 23/1; 62/2; 75/2; 117/2; 163/1; 277/3; 319/3; 363/1; 413/3; 443/3

Query Match 14.8%; Score 306.5; DB 2; Length 499;
 Best Local Similarity 24.5%; Pred. No. 5.2e-17;
 Matches 80; Conservative 72; Mismatches 134; Indels 41; Gaps 7;

```

Oy      79 LPKSSILGGQFAIWRNGPPOERSRLCSIALSGMLGCFATLLIGFTISLT----- 132
Db      166 IPGGSVLIS-----WPLSEKSTAMALFTTGNOIG-----IAMSFLAKLCQLHFE 214
Oy      132 GMPFVYITFGVGCVCCLLMFYVITYDDPVSYPISTSEKEYIISLKQOVGSSKQPLPIK 191
Db      215 GMPLVYIVGLIGAVPLVIMHVRADKPRESKYITATELTYI-----KGGQRNRRAE 267
Oy      192 AMRSLP-----TMSICLGCPSHQLVSTMTVVIPTIYSSVYHVNIRNDGLSALP 242
Db      268 TIVRATPYMKIILNGCVCAICAFQSFQSFVLAVALYLPKYNQIAFKMNLTHNGIWSLP 327
Oy      243 FIYAVVIGVGGYLAFLTKKFRLITVRKIATTLIGSLPSNALIYSLPLNNGYITATL 302
Db      328 FFIQMTKLFLPAIADKVKQRKNATVTKVSNAISFSAITVIAA---GPFDSAEI 384
Oy      303 LPLSCGLSTLCSG---GIYINVLADIPRYSFLMGASRGFSSIAPIVPTVSGFLLSOD 358
Db      385 VQLSIVSMAASAVPYGNTSIVYVAPQFTAFISSYQALQIASTAPIYIGRTISHG 444
Oy      359 PERGMRNVFLLFAVNLGLLFYLIIG 385
Db      445 TIYEMKCAFYSLAGVLAATGLIFQIRG 471

```

RESULT 12

729418
 hypothetical protein F21F8.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29418
 R:Wilson, R.; Favello, A.; Le, T.T.
 submitted to the EMBL Data Library, April 1997
 A:Description: The sequence of C. elegans cosmid F21F8.
 A:Reference number: 220618
 A:Accession: T29418
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-530 <MIL>
 A:Cross-references: EMBL:097000; PIDN:AA04797.1; GSPDB:GN00023; CESP:F21F8.11
 A:Experimental source: strain Bristol N2; clone F21F8
 C:Genetics:
 A:Gene: CESP:F21F8.11
 A:Map position: 5
 A:introns: 16/1; 64/2; 112/3; 344/3; 417/1; 452/1

Query Match 14.7%; Score 304.5; DB 2; Length 530;
 Best Local Similarity 22.9%; Pred. No. 8e-17;
 Matches 110; Conservative 69; Mismatches 186; Indels 115; Gaps 15;

```

Oy      15 LGSARYGIALVTH-FCNFTTAQNYIIN-----ITMVAMVNSTSPQSOLNDS 60
Db      38 MCOMFAHIGLSTSCMCNSTAVA---LMTNNTATLVGTESTILMSILEKNTSESQELGG 94

```

```

QY      61 S-----EVLPSVDSGG-----72
      95 DDGESCRLKBEKVIKDIQGTPIWVSOGYIVSAFLGGLFIPSPAGVLDVDFSRHILS 154
QY      72 -----LSKAPKSLPAKSSILG--GQFA-----IMERGPPQERSRLC 106
      155 VAILMLTATLMLPLVSLIYCEKGAFAKRFVWGISEMLPISINSMTKMLPINKSLAA 214
QY      107 SIALSG-WLCCGTAAILIGGFISETLGPVFYIFGVCYCCCLMFVYIYDDPVSYIWI 165
      215 SVFTAGNLSGMPFNILVAELCASSFGWSSIFYSASLFGISLMVLMHLTVRNSPHNTMI 274
QY      166 STEKEYIISLSKOQVSS--KQPLPKAMLSLPIWISICLGFSGHQLVSTWVYIPIYI 224
      275 HKRELDIYANNIPKHPSSVYKKTPIWMLTSKVPWMSLMPNSVMGNMIALIPIYIPIYF 334
QY      225 SSYVHVNIKRDNGLLSALPFIYAWYIGVGYLADFLTKRF---RLITVKRIATILGSL 280
      335 KDVIMLMDVQSNFGYSAIPHISNLIAKILWIGMDKMRKILSPSAYVKLSQFASMG-- 393
QY      281 PSSALIVSLPLNSGYIAT---ALITLSC-----GLSTLCOSGIYINVLDIAPRYSFL 332
      393 -ISVSCFELRYMN---CATPFYALVLLSVSAFEGLSI---SGFYTSLISAPSHIGTL 444
QY      333 MGASRGFSSIAPIVPIYVSGFLLSODPEFGMRNPFELFAVNILGLLYLIFGEGADVQEW 392
      445 TSLATVIGVGRMFTPLMISTYKTYGTAEBMGHULLIYVPASATGCTIFLFGSGDVQNM 504

```

RESULT 13

```

T29968
hypothetical protein ZK682.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29968
R:Du, Z.; Le, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid ZK682.
A:Reference number: 220714
A:Accession: T29968
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-506 <DU2>
A:Cross-references: EMBL:U01110; PIDN:AAA82415.1; CESP:ZK682.2
A:Gene: CESP:ZK682.2
A:Introns: 37/3; 82/3; 163/3; 314/3; 359/1; 469/1

```

```

Query Match      13.6%; Score 282.5; DB 2; Length 506;
Best Local Similarity 24.6%; Pred. No. 4.3e-15;
Matches 82; Conservative 70; Mismatches 140; Indels 41; Gaps 11;

QY      85 ILG-GQFAIW-----ERMGPPQERSRLCSIALSGMLLGFATILIGFISETL-----132
      158 VMFGQGVLMPCMLVLIAQWFPVNEKSTALAIATGNQL---SVIAMPATAEICQJLPW 213
QY      132 GMPFVYIFGVCYCCCLMFVYIYDDPV-STPWISTEKEYIISLSKOQVSSKQPLPI 190
      214 GMPMAFHYAVCGIYWCILWYMYVSDPCHADKLSRDELHIITF---ERVLMRPOHNM 270
QY      191 KAMRLSLIWSICLGFSGHQLVSTWVYIPIYIISVYHVNIKRDNGLLSALPFIYAWYIG 250
      271 MALMKSPPVWMLAASSFAMHNTVGTITLYLPIYKTVLNMSTLSGMLSALPFIQLLSK 330
QY      251 MVGGLADFLTKRFLITVKRIATILGSLPSSALIVSLPYL-----NSGYTATALLT 304
      331 VFYAGMAF--SARKRDWMDINKRITFCNSASFGIAICFGLLCLDCQORG-----AAIF 383
QY      305 LSCGLSTLCQSGIYI-----NVLDIAPRYSFLMGASRGFSSIAPIVPIYVSGFLSDP 359
      384 LIC--LAMCFVSGYIPGYNTSAVTIAPGOTAAIAAFSRFGQIASSVAPYHIGAVTKOGT 441

```

```

QY      360 EFGMRNPFELFAVNILGLLYLIFGEGADVQEW 392
      442 ADEWKIVFAVIAICVGVIGIFQCGGTASLDWM 474

```

RESULT 14

S44742

C02C2.4 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 09-Sep-1997

C:Accession: S44742

R:Wilson, R.

submitted to the EMBL Data Library, September 1993

A:Description: Sequence of the C. elegans cosmid C02C2.

A:Reference number: S44737

A:Accession: S44742

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-568 <WIL>

A:Cross-references: EMBL:L23649; NID:9388566; PID:9388572

C:Genetics: A:Introns: 39/1; 55/3; 103/3; 128/3; 206/3; 266/1; 310/3; 361/3; 406/1; 436/2; 487/3

```

Query Match      13.5%; Score 280; DB 2; Length 568;
Best Local Similarity 22.8%; Pred. No. 7.8e-15;
Matches 92; Conservative 68; Mismatches 191; Indels 52; Gaps 10;

```

```

QY      5 ETLIPKVPYSLCSARVGIATVL-----HFCNFTTIAONVIMNTWVAMVNSTSPQSOLN 58
      161 ERFSPRHVLQISVALYILVTVITPFLATHFGYFVFLARIQM-----203
QY      59 DSSEYLPVDSFGGLSKAPKSLPAKSSILGQPAIEMKRGPPQERSRLCST-ALSGMLIGC 117
      203 -----GLGEG-FVEPTNNAITG-----NMFPSEKSTALSIPTLGNQISA 242
QY      118 FTAILIGFISETLGMPVVFYIFGVCYCCCLMFVYIYDDPVSYPMWISTEKEYIISL 177
      243 AGSPWAAVACASDLGWPPTFFAGIFATGWSLTFEFTASSHPAKVMMTKREKEYLLANV 302
QY      178 KOQVSSK--QPLPKAMLSLPIWISICLGC-FSHQWLVTWVYIPIYIISVYHVNIKRD 234
      303 VKKVKSEKTRSIPIYSKILTS-PAFLGQLCHFFVNLPMFLFOYLIPSEYFVHLGVIA 361
QY      235 NGLSALPFIYAWYIGVGYLADFLTKRRLIT-VKIKATILGSLPSSALIVSLPLN 293
      362 NGFTALPNIIFNMIFKVVWVGIGIDLKENKILSNKRAKAVSHGVAASFSSLSLILAFV 421
QY      294 SGYIATALLTSCGLSTL--CQSGIYINVLDIAPRYSFLMGASRGFSSIAPIVPIYVS 351
      422 DCSNPTTGLIIFCLMYSMGTFVSGFYTSLSLAPQYIATYSAISMFPYAMIGRLTTPAVM 481
QY      352 GFLLSQDPEFGMRNPFELFAVNILGLLYLIFGEGADVQEMAK 394
      482 SMFRKDTAAEMQNIIFGICSLAHIFSGSIFLLFGSGELQDMAK 524

```

RESULT 15

ZK652.10 protein - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 14-Sep-1994 #sequence_revision 12-May-1995 #text_change 09-Sep-1997

C:Accession: S44900

R:Du, Z.

submitted to the EMBL Data Library, May 1993

A:Description: Sequence of the C. elegans cosmid ZK652.

A:Reference number: S44613

A:Accession: S44900

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-420 <DU2>

A:Cross-references: EMBL:L14429; NID:9289765; PID:9289766

C:Genetics:
A;introns: 60/1; 104/3; 155/3; 200/1; 230/2; 281/3; 337/1; 354/3; 379/1

Query Match 13.5%; Score 279.5; DB 2; Length 420;
Best Local Similarity 24.5%; Pred. No. 6e-15;
Matches 79; Conservative 62; Mismatches 166; Indels 15; Gaps 7;

```

QY 80 PAKSSILGQFAIWERNGPQERSRLCSI-ALSGMLGCTAILIGFISETIGMPVEFY 138
   | : : : | | : : | : : | : : | : : | : : | : : | : : | : : |
Db 5 PNNNAIIG-----NMFPSSSEKSTALSIPTIGNQIASAGSPMVAACASDLGMPATFY 57

QY 139 IFGVCVCVCCLLMFVVIYDDPVSYPMISTSEKEYIISLKOVGSSK--OPLPIKAMLR 196
   | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 58 FAGIFATGNSILMFPTASSHPAKVKMMTKREKEYILANVKKYHKSEKTRSIPIYSKILTS 117

QY 197 LPIWISICLGC-FSHQMLVSTMVYIPTIYISSVYHVNIRDNGLSALPEYIVAWYIGWVGY 255
   | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 118 -PAFLGOLCHFEFVNLFMTLFOIYLPSEYFEVHLGLVIANGTFTAJPNIEFMIFKVVWGI 176

QY 256 LADFLTKKFRLLT-VKRTATILGSLPSSALIVSLPYLNSGYTTATALLTLSCGLSTL-- 313
   | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 177 GIDKLEKNKILSNTKAVKSHGVAISGSSSFLILAEFYDCSNPTGLIFFCLMYSSMGT 236

QY 313 COSGIYINVLDIAPRYSSFLMGASRGFSSIAPIVPTVSGFLLSQDEFGMRNVFFLLFA 372
   | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 237 FVSGFTSLSLAPQYTAIRSAISMFVAMIGRLTTPAVMSMERKDGTAEMONIFIGCSL 296

QY 373 VNLGLLFYLIAGEADVOEWAK 394
   | : : : | : : | : : | : : | : : | : : | : : | : : | : : |
Db 297 AHIFSGSIFLLFGSGELQDWAK 318

```

Search completed: May 25, 2000, 14:27:17
job time: 2304 sec

THIS PAGE BLANK (USPTO)

OY	61	SEVL	PDSFGGSLKAKSPKSPAKSIIIGGQFATMERNGPOERSRCLSI	120
DB	61	SEVL	PLPDSFGGSLKAKSPKSPAKSIIIGGQFATMERNGPOERSRCLSI	120
OY	121	ILIGG	FISFTLGMPEVYIFGGVGCVCCLLMFVIVYDDPVSYPMWISTSEK	180
DB	121	ILIGG	FISFTLGMPEVYIFGGVGCVCCLLMFVIVYDDPVSYPMWISTSEK	180
OY	181	VSSSK	PLPKAMLRSLPTWISTCLGCFESHOMVSTWVVIYPIYSSVYHNRDNL	240
DB	181	VSSSK	PLPKAMLRSLPTWISTCLGCFESHOMVSTWVVIYPIYSSVYHNRDNL	240
OY	241	LEPI	YAMVIGMVGVLADLTLTKKFLIVRKATIIIGSLPSALLVSPYLSNG	300
DB	241	LEPI	YAMVIGMVGVLADLTLTKKFLIVRKATIIIGSLPSALLVSPYLSNG	300
OY	301	ALLT	SCGSLTLCQSGITINVLDIAPRYSFTLMGASRGFSSTAPVIVPVSG	360
DB	301	ALLT	SCGSLTLCQSGITINVLDIAPRYSFTLMGASRGFSSTAPVIVPVSG	360
OY	361	FGMR	NYFLLFVNLGLFLYILFGADVQENAKKKRLRL	401
DB	361	FGMR	NYFLLFVNLGLFLYILFGADVQENAKKKRLRL	401
RESULT	2			
ID	NPT1_HUMAN	STANDARD;	PRT;	465 AA.
AC	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	15-OUL-1998	(Rel. 36, Last annotation update)		
DE	RENAL SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1 (SODIUM/PHOSPHATE COTRANSPORTER 1) (NA(+)/PI COTRANSPORTER 1) (RENAL SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1) (RENAL NA(+)-DEPENDENT PHOSPHATE COTRANSPORTER 1) (NA/PI-4).			
DE	SLC17A1 OR NPT1.			
GN	Slc17a1			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
NC	[1]			
NP	SEQUENCE FROM N.A.			
RP	TISSUE-KIDNEY.			
RC	MEDLINE: 94117004.			
RX	Chong S.S., Kristjansson K., Zoghbi H.Y., Hughes M.R.;			
RA	"Molecular cloning of the cDNA encoding a human renal sodium phosphate transport protein and its assignment to chromosome 6p21.3-p23."			
RT	Genomics 18:355-359(1993).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-KIDNEY CORTEX;			
RX	MEDLINE: 95126933.			
RA	Miyamoto K.-I., Tatsumi S., Sonoda T., Yamamoto H., Minami H.,			
RA	Taketani Y., Takeda E.;			
RT	"Cloning and functional expression of a Na(+)-dependent phosphate co-transporter from human kidney: cDNA cloning and functional expression."			
RL	Biochem. J. 305:81-85(1995).			
CC	-1- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY. MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.			
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.			
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN KIDNEY CORTEX, LIVER AND BRAIN BUT NOT IN OTHER TISSUES.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			

[illegible]

```

GN SLIC17A1 OR NP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE: 97021433.
RA Li H., Xie Z.;
RT "Molecular cloning of two rat Na+/Pi cotransporters: evidence for
RT differential tissue expression of transcripts."
RL Cell. Mol. Biol. Res. 41:451-460(1995).
CC -I- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U28504; AAC52487.1; -
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79
FT TRANSMEM 99
FT TRANSMEM 117
FT TRANSMEM 137
FT TRANSMEM 176
FT TRANSMEM 196
FT TRANSMEM 219
FT TRANSMEM 219
FT TRANSMEM 260
FT TRANSMEM 280
FT TRANSMEM 299
FT TRANSMEM 319
FT TRANSMEM 337
FT TRANSMEM 357
FT TRANSMEM 363
FT TRANSMEM 383
FT TRANSMEM 399
FT TRANSMEM 419
FT TRANSMEM 431
FT TRANSMEM 451
FT CARBOHYD 39
FT CARBOHYD 47
FT CARBOHYD 56
FT CARBOHYD 56
SQ SEQUENCE 465 AA; 51350 MM; 28BBB8DC5C0AC52 CRC64;

Query Match 46.0%; Score 954; DB 1; Length 465;
Best Local Similarity 41.5%; Pred. No. 6,5e-65;
Matches 193; Conservative 79; Mismatches 127; Indels 66; Gaps 6;

QY 3 VDETLIPRKVPSICSAARYGIALVLFHCNFTTIAQNTVIMNTTAMVNSTSPQSOLNDS-S 61
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MEMRCIPKRVGFCSPFRYGIALILHFCNIVYMAQRCVCLNTJTMAMVNTKTPPHLSNKSVA 60
QY 62 EVL-----PVDSF-----GGLSKA----- 76
   ||| |||
DB 61 EMDLNVKNPVHWSMDIOGLVLSVFLGVNVIVQPVGLSSVFLGAPMEKIISSLSFLSVLS 120
QY 76 ----PKSLP-----AKSSILGGQFAIMERNKPPQERSLCLIALSGMLGC 117
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 LILPPAQAQCALVYICRVLQGIAGQAVSTGQHGIMVKNAPLERGLTJMTLSGFWGP 180
QY 118 FTAILIGFISLTPMPFEVYIGFGVGVCCCLMFVYIDDPVSPYPISTSEKEYITSSL 177
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 FIALIVSGFICDLGWMVMYIIFGIVGVLSLFWFILLFPDPNNHNPMSSEKDYITSSL 240
QY 178 KOQVSSKQPLPIKAMLRSLPIWSICIGCFSHQMLVSTWVVIPIYITSSVYHVNIRNGL 237
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 MOQVHSGROSPLIKAMKSLPLMAIILNSFAFTSMNNLVTYPTFISTLTHVAVRENGL 300
QY 238 LSLPPIVAMVIGMGVYLADELTKK-FLIIVRKIATITLSLPSALLVSPYLSNGY 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 LSLPPIVAMVIGMGVYLADELTKK-FLIIVRKIATITLSLPSALLVSPYLSNGY 236
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 297 IFATALLTSCGLSTCGSGIYINVLDIAPRYSSFLMGASGRGSSIAPIVTPVSGFLS 366
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

DB 361 YSTVIFLLANSTLSFSCQQLINADLADIRYGGFLKAVTALIGFGLISSTLAGLIIN 420
QY 357 QDEPFGKRVNFFLLFAVNLGLFLYIFGENDVQWAKKRLRL 401
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 QDEYAWHKNFFLMAGINVTCLAFYLLFAGKDIDQWAKKRTTL 465

RESULT 4
NP1_MOUSE
ID NP1_MOUSE STANDARD; PRT; 465 AA.
AC 061983;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RENAL SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1 (SODIUM/PHOSPHATE
DE COTRANSPORTER 1) (NA+)/PI COTRANSPORTER 1) (RENAL SODIUM-PHOSPHATE
DE TRANSPORT PROTEIN 1) (RENAL NA+-DEPENDENT PHOSPHATE COTRANSPORTER 1).
GN SLIC17A1 OR NP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RX MEDLINE: 95335846.
RA Chong S.S., Korzak C.A., Liu L., Kristiansson K., Dunn S.T.,
RA Boudreau J.E., Hughes M.R.;
RT "Cloning, genetic mapping, and expression analysis of a mouse renal
RL sodium-dependent phosphate cotransporter."
CC Am. J. Physiol. 268:F1038-F1045(1995).
CC -I- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -I- TISSUE SPECIFICITY: KIDNEY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77241; CAA54459.1; -
KW MGD; MGI:103209; SLIC17A1.
KW Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
FT TRANSMEM 79
FT TRANSMEM 99
FT TRANSMEM 117
FT TRANSMEM 137
FT TRANSMEM 176
FT TRANSMEM 196
FT TRANSMEM 199
FT TRANSMEM 219
FT TRANSMEM 260
FT TRANSMEM 280
FT TRANSMEM 304
FT TRANSMEM 324
FT TRANSMEM 337
FT TRANSMEM 356
FT TRANSMEM 363
FT TRANSMEM 383
FT TRANSMEM 399
FT TRANSMEM 419
FT TRANSMEM 429
FT TRANSMEM 449
FT CARBOHYD 39
FT CARBOHYD 47
FT CARBOHYD 47
FT CARBOHYD 56
FT CARBOHYD 56
SQ SEQUENCE 465 AA; 51589 MM; C67BE25A2C291EEF CRC64;

Query Match 45.6%; Score 946; DB 1; Length 465;
Best Local Similarity 40.2%; Pred. No. 2,6e-64;
Matches 187; Conservative 78; Mismatches 134; Indels 66; Gaps 3;

QY 3 VDETLIPRKVPSICSAARYGIALVLFHCNFTTIAQNTVIMNTTAMVNSTSPQSOLNDS-- 61
   ::::: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1 MEMRCIPKRVGFCSPFRYGIALILHFCNIVYMAQRCVCLNTJTMAMVNTKTPPHLSNKSVA 60
QY 61 -----SEVLPVDSFGGLSKAPSLP- 81
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 61 EMLDNVKNVYNSPDIQGLILSSVEFGMVVQAPVGLISGLIYPMKRIIGSSILFSLSLMS 120
OY 81 -----AKSIIIGGFAIWERNGPQPERRLCSIALSGMLGC 117
Db 121 LLIPPAOYGAALVYCRVLOGIAGCTVSTGHEIWMKAPLERGLTSMILSGVMP 180
OY 118 FPAIIGFISFTLGMPEVYIFGVGCYCLLMFVYIDDPVSYWISTSEKEYITSSL 177
Db 181 FVLLVSGFICDLGMPWFYIFGIYGVYLSWFLFEDDDKDHMYSSSEKDYITSSL 240
OY 178 KOOVSSKQPLPIKMLRSLPIMSICLGFSGHOMLVSTWVVIPIYISSYHVNRDGL 237
Db 241 MOOASSGRSLPIKMLKSLPMLAILNSFAFMSLSLVTPPFIISTVLHVNRDGL 300
OY 238 LSALPIYAWVIGWVGLADFLTRK-FRLITVRKATILSLSPSSALIVSLPYNSGY 296
Db 301 LSLPYLAVIGIILAGOMSDFLRKISYTVKRLFTTLSPFCVPIYIMCLYLHSTNF 360
OY 297 ITATALLTSLGSLTSCSGIYINVLADIPRYSSFLMGASRGSSIAPIVPTVSGFLS 356
Db 361 YSTVIFELIANSTLSFSYCGQLINLADIAPRYGFLKAVTALIGMGGLISSTLGLILN 420
OY 357 GDEPFGMRNVFLLFAVNLGLFLIFLEADVOEMAKERKILRL 401
Db 421 QDPESWFKIFLMSIINVIVIFLFAKGEIDOMAKERKILRL 465

```

RESULT 5

NPT3_RABIT STANDARD; PRT; 465 AA.

```

AC 028722.
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE RENAL SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 1 (SODIUM/PHOSPHATE
DE COTRANSPORTER 1) (NA(+)/PI COTRANSPORTER 1) (RENAL SODIUM/PHOSPHATE
DE TRANSPORT PROTEIN 1) (RENAL NA+-DEPENDENT PHOSPHATE COTRANSPORTER 1)
DE (NAP-1).
GN SLC17A1 OR NPT1.
OS Erythrocytes cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-KIDNEY CORTEX;
RX MEDLINE; 92052140.
RA Werner A., Moore M.L., Mantei N., Biber J., Semenza G., Murer H.;
RT "Cloning and expression of cDNA for a Na/PI cotransport system of
RT kidney cortex."
RL Proc. Natl. Acad. Sci. U.S.A. 88:9608-9612(1991).
CC -1- FUNCTION: IMPORTANT FOR THE RESORPTION OF PHOSPHATE BY THE KIDNEY.
CC MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE INTO CELLS VIA
CC NA+ COTRANSPORT IN THE RENAL BRUSH BORDER MEMBRANE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: KIDNEY CORTEX AND LIVER.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M76466; AAA31461.1;
CC
CC PFAM: pf00083; sugar_tr; 1.
CC
CC Transport; Transmembrane; Glycoprotein; Symport; Sodium transport.
CC
CC TRANSMEM 79 99
CC TRANSMEM 109 129
CC TRANSMEM 171 191
CC TRANSMEM 198 218
CC TRANSMEM 255 275
CC TRANSMEM 304 324

```

```

FT TRANSMEM 337 357 POTENTIAL.
FT TRANSMEM 363 383 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 428 448 POTENTIAL.
FT CARBOHYD 47 47 POTENTIAL.
FT CARBOHYD 56 56 POTENTIAL.
SQ SEQUENCE 465 AA; 51798 MW; 29B33FB1EADDCBAD CRC64.

```

Query Match 44.9%; Score 932; DB 1; Length 465;
 Best Local Similarity 39.4%; Pred. No. 2.9e-63;
 Matches 183; Conservative 77; Mismatches 139; Indels 66; Gaps 3;

```

OY 3 VDETLPRKVPSELCAKRYALVLRHCFNTTIAQVANIITVAVANSTSPQSOLNDSE 62
Db 1 MDNPPSRKPCFCSEFRVLALEMHRCNIVITIAORCCSLTVAVANNITNLGSPNTSAE 60
OY 63 -----VLPVDSFGLSKAPKSLP----- 81
Db 61 KRLDMTKNPVYMWSPDVQGIIFSSIFYGAFLIQIVGYISGLITKILGFLFLSLVS 120
OY 81 -----AKSIIIGGFAIWERNGPQPERRLCSIALSGMLGC 117
Db 121 IFIPQAAVGETWIIYCRVQGITQGTVTTAQHEIWMKAPLERGLTSMILSGFLILGP 180
OY 118 FPAIIGFISFTLGMPEVYIFGVGCYCLLMFVYIDDPVSPWISTSEKEYITSSL 177
Db 181 FVLLVVTGICISLGMPEVYIFGVGCYCLLMFVYIDDPVSPWISTSEKEYITSSL 240
OY 178 KOOVSSKQPLPIKMLRSLPIMSICLGFSGHOMLVSTWVVIPIYISSYHVNRDGL 237
Db 241 IQGSSSTRQSLPIKMLKSLPMLAILNSFAFMSLSLVTPPFIISTVLHVNRDGL 300
OY 238 LSALPIYAWVIGWVGLADFLTRKFRIT-VKRIATILGSLSPSSALIVSLPYNSGY 296
Db 301 LSLPYLAVIGIILAGVYAGTADFLMSRNMISLAIKRLFTALIGLPIYFSCMLYLSSGF 360
OY 297 ITATALLTSLGSLTSCSGIYINVLADIPRYSSFLMGASRGSSIAPIVPTVSGFLS 356
Db 361 YSTVIFELIANSTLSFSYCGALINLADIAPRYGFLKAVTALIGMGGLISSTVAGFLS 420
OY 357 GDEPFGMRNVFLLFAVNLGLFLIFLEADVOEMAKERKILRL 401
Db 421 QDPESWFKIFLMSIINVIVIFLFAKGEIDOMAKERKILRL 465

```

RESULT 6

NPT3_HUMAN STANDARD; PRT; 436 AA.

```

AC 000624.
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE SODIUM-DEPENDENT PHOSPHATE TRANSPORT PROTEIN 3 (SODIUM/PHOSPHATE
DE COTRANSPORTER 3) (NA(+)/PI COTRANSPORTER 3).
GN SLC17A3 OR NPT3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA Ruddy D.A., Krommal G.S., Lee V.K., Mintler G.A., Quintana L.,
RA Domingo R., Meyer N.C., Basava A., McClelland E., Fullan A.,
RA Mapa F.A., Moore T., Thomas W., Loeb D.B., Harmon C., Tsuchihashi Z.,
RA Wolff R.K., Schatzman R.C., Feder J.N.;
RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: MAY BE INVOLVED IN ACTIVELY TRANSPORTING PHOSPHATE
CC INTO CELLS VIA NA+ COTRANSPORT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: 091328; AAB82085.1; -
DR EMBL: 090544; AAB53422.1; -
KW Transport; Transmembrane; Glycoprotein; Phosphorylation; Symport;
KW Sodium transport.
FT TRANSMEM 26 46 POTENTIAL.
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 211 231 POTENTIAL.
FT TRANSMEM 273 293 POTENTIAL.
FT TRANSMEM 317 337 POTENTIAL.
FT TRANSMEM 366 406 POTENTIAL.
FT CARBOHYD 47 47 POTENTIAL.
FT CARBOHYD 56 56 POTENTIAL.
FT CARBOHYD 68 68 POTENTIAL.
FT CARBOHYD 69 69 POTENTIAL.
SQ SEQUENCE 436 AA; 47350 MW; DF02F618E83A572F CRC64;

Query Match 35.8%; Score 744; DB 1; Length 436;
Best Local Similarity 38.5%; Pred. No. 4e-49;
Matches 162; Conservative 56; Mismatches 107; Indels 96; Gaps 5;

OY 10 RAYPSICARVGIATLVHFCNFTTIAQNYIMNTWAMVNSTSPQSLDSEVLDPYDF 69
DB 8 RKPDPDCSLRYGLALIMHRSNFTMITQRYVSLAIATIAMVTTQOGLSNASTEGPVADAF 67
OY 70 GGLSKRPKSLPAKSSI-----LG----- 88
DB 68 NNSISIKKEPDKASVYOWSPETQGIIFSSINYGILTLIPSGYLAGIFGAKKMLGAGLL 127
OY 88 -----GQAFIMRWGPPOEBSRLCSTALS 111
DB 128 ISSLLTFPLPADFGVILVIMRYQGAQMAQMGQFTIMAKMPLPERSKLTITAGS 187
OY 112 GMLGCFATLIGCFISETLGWPFYIFGVCVCCLLMFVYIDDPVSYPMISTSEKE 171
DB 188 GSAFGSFIILCVGLISQALSMWPFIFYIFGSCVCCLLMFVYIDDPMHHCISVREKE 247
OY 172 YIISSLKQGVSSKOPLPKAMRLPIWISICGSCSHQWLVSTMVYIPTYISSYTHN 221
DB 248 HILSLAQOPSSPGRAVPKAMVTCPLVAIFLGFFSHFWLCTIILTYIPTIISTLLHN 307
OY 232 IRDNGILSLALPFIYAVNIGWVGYLADFLTKK-FRLITYRKIATYI-----L 277
DB 308 IRDSVLSLPFIYAAASCIILGGLADFLSLRLALRIYRKLFSSLDQVSSWESQGL 367
OY 278 GSPSSALIVSLPYLNSGYTTATATLTLSCGLTSCQSGIYINVLDIAPRYSSFLMGASR 337
DB 368 GSSQESSLPLDLDSSSVRLISLVGMSFCLQSTCLAMSFTRLD-----KQNFKIGPKR 423
OY 338 G 338
DB 424 G 424

RESULT 7
Y006_CAEEL STANDARD: PRT: 576 AA.
AC P34644;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEICAL 52.2 KD PROTEIN ZK512.6 IN CHROMOSOME III.
GN ZK512.6
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Peloderidae; Caenorhabditis.

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE: 94150718.

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Dublin R., Favell A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
RA Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA Watson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: 222177; CAAB0150.1; -
DR PIR: S40767; S40767.
KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
KW Sodium transport.
FT TRANSMEM 70 90 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 151 171 POTENTIAL.
FT TRANSMEM 175 195 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 240 260 POTENTIAL.
FT TRANSMEM 296 316 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 382 402 POTENTIAL.
FT TRANSMEM 407 427 POTENTIAL.
FT TRANSMEM 438 458 POTENTIAL.
FT TRANSMEM 473 493 POTENTIAL.
SQ SEQUENCE 576 AA; 63108 MW; 71F3A9EFBE5F84CC CRC64;

Query Match 22.5%; Score 466.5; DB 1; Length 576;
Best Local Similarity 32.3%; Pred. No. 4.3e-28;
Matches 101; Conservative 65; Mismatches 134; Indels 13; Gaps 6;

OY 92 IWRMGPPPOEBSRLCSTALSGMLGCFATLIGCFISETLGWPFYIFGVCVCCLLM 151
DB 198 VWRVAPRPMERSKLTATTAFTAGTAGAVLGPLSAFLVSYWMAAPFLVGVGYNATIM 257
OY 152 FVYIDDPVSYPMISTSEKEKYSIKQOYG--SKOP----LPKAMRLPIWISICLG 205
DB 258 FCVTFEKPAPHFPTISOEKEKIFI---EDAIGHVSNHPTIRSIPIKAIYVSKVYMAIYA 313
OY 206 CFSHOWLVSTMVYIPTYISSYTHVNRDGLSALPFIYAVNIGWVGYLADFLTKK 265
DB 314 NFARSWTFYLLONOLTYKMEALGMIDGSLAIPHLVWGVVGLGGLADYLSNKKI 373
OY 266 RLIT-VRKIATILGSLPSSALIVSLPYLNSGYTTATATLTLSCGLTSCQSGIYINVDI 324
DB 374 LSTTAIRKPIFGCGFGEAFLVATYTSDD-TTATMALAAVMSGFAISGFNVNHLDI 432
OY 325 APRYSFSLMGASRGFSSIAVIVPTVSGFLSDPFGWRNVFLLFAVNLGLLFYLI 384
DB 433 APRYATILMGFSNGICTLGLTCTPVTLEAPTANS-KHGWTSVPLASLIHFTGVTFYAVY 491

OY 385 GADVOEMAKERK 397
 DB 492 ASGLOEMAEKE 504

RESULT 8
 YL2D_CAEEL STANDARD: PRT: 493 AA.
 AC 003567;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 52.7 KD PROTEIN C38C10.2 IN CHROMOSOME III.
 GN C38C10.2.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favell A., Fraser A.,
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 RA Johnston L., Jones M., Kersey J., Kirsten J., Lalster N.,
 RA Johnstone P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Shownkeen R.,
 RA Sims M., Smaildon N., Smith A., Smith M., Sonhammer E., Staden R.,
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 RA Watson R., Watson A., Wellstock L., Wilkinson-Spoat J.,
 RA Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP REVISIONS.
 RA Jones S.J.M.;
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z19153; CAAT9549.1; -;
 DR PIR: S28286; S28286.
 DR NORMP: C38C10.2; CE086647.
 KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
 KM Sodium transport.
 FT TRANSMEM 10 30 POTENTIAL.
 FT TRANSMEM 85 105 POTENTIAL.
 FT TRANSMEM 112 132 POTENTIAL.
 FT TRANSMEM 144 164 POTENTIAL.
 FT TRANSMEM 175 195 POTENTIAL.
 FT TRANSMEM 205 225 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 FT TRANSMEM 311 331 POTENTIAL.
 FT TRANSMEM 348 368 POTENTIAL.
 FT TRANSMEM 375 395 POTENTIAL.
 FT TRANSMEM 406 426 POTENTIAL.
 FT TRANSMEM 441 461 POTENTIAL.
 FT TRANSMEM 493 AA; 52737 MW; 60792715D32553DD CRC64;
 SQ SEQUENCE

Query Match 22.1%; Score 459.5; DB 1; Length 493;
 Best Local Similarity 25.7%; Pred. NO. 1.2e-27;
 Matches 121; Conservative 82; Mismatches 181; Indels 87; Gaps 11;

OY 6 TLPRKVPSCARGYATLVHFCNFTIAQNVIMNTVAVANSTSPQSLDSS-----62
 DB 5 TTKRLVP---STRFALSLVFGCCLVTYMMRTNMSFAVCNENKTDGVKVSRCGK 61
 OY 62 EVLPVDSFGGL-----SKAPSLPAK-----SSLGGQTA-----92
 DB 62 EMFVFNSSSVYIGFDMDKQTTGMYLSSFFYGYSQIIIGHLASRYGRVFTVLGS 121
 OY 92 -----IWRMGPOERSRLCSIALSGM 113
 DB 122 ALLTLNLPVARTSEVALAILRAIGLQGTPEPMHTNMSWCPPELSVLGCVYTAGA 181
 OY 114 LIGFTAILIGFISE---TLGMPFVYIFGVCVCCLLMFVYIDDPVSPWISTSEK 170
 DB 182 QIGNVYIPLSGFLCEFGFDGMSIFPIIGVFCVLTAVWVYSSDKPAPHTPEEK 241
 OY 171 EYIISLKKQOVSSKQPLP---IKAMRLSPINSICLGCSSHOMLVSTMYVYIPTYIS 226
 DB 242 QYIVAVASMGKDTGKVPSTPWIK-ILTSPAVNACVAGHFGAGVATMLVSLPSFLKD 300
 OY 227 YHVHVRNGLLSALPEIVANVIGWVGXYLADPELLTKK-FRLITVRKATILGSLPSAL 285
 DB 301 VGLNLSLGAVASIPYIATVLAIVAGVLDTLRSKGLISTLTRRAMVALLIGGIF 360
 OY 286 IYSLPINSGYTTATATALLTSCG-LSTLCOSGIYINVDIAPRYSSFLMGASRGFSIA 343
 DB 361 LVASGTCGGQ-DVVIIFITCGMAISGQYAFVNVLEIAPPSGVTGNTISMA 419
 OY 344 PVIYVPSGFLSLQPEGRVNFVFLFRVNLGLLFLVGEADVQSMK 394
 DB 420 GLISPAVSSYLPNGTOEBMOMVLMITAGILITGALLSIFASGEVQPMK 470

RESULT 9
 YRT3_CAEEL STANDARD: PRT: 544 AA.
 AC 010046;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE HYPOTHEICAL 60.5 KD PROTEIN T07A5.3 IN CHROMOSOME III.
 GN T07A5.3.
 OS Caenorhabditis elegans.
 CC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Buck D.;
 RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE COTRANSPORTER 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z48055; CAAB8134.1; -;
 DR NORMP: T07A5.3; CE01648.
 KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
 KM Sodium transport.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 105 125 POTENTIAL.
 FT TRANSMEM 128 148 POTENTIAL.
 FT TRANSMEM 155 175 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 221 241 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.

FT TRANSMEM 324 344 POTENTIAL.
 FT TRANSMEM 381 381 POTENTIAL.
 FT TRANSMEM 385 405 POTENTIAL.
 FT TRANSMEM 417 437 POTENTIAL.
 FT TRANSMEM 451 471 POTENTIAL.
 SQ SEQUENCE 544 AA; 60543 MW; 4615B3C27D9F86CC CRC64;

Query Match 20.9%; Score 434; DB 1; Length 544;
 Best Local Similarity 25.4%; Pred. No. 1,1e-25;
 Matches 119; Conservative 85; Mismatches 167; Indels 98; Gaps 13;

5 ETLIPKRVPSLCARVGIALLVLF-----CNFTTIANQYINIMITAMVNSTSPQS 55
 33 QTTTEKKKFLKRVKQIALAHFPAISGINSNF-GVAKNMVN-----NFTDYG 84
 56 QLND-----SSEVLVPVDFSGISKAPKSLPAKSSILGQFA----- 92
 85 EYHEREFMTGAEVGMWESSFFYGVAASQIPA--GVLAKEFAFNKIFMLGILVASEFNNIL 142
 92 -----IWERGPPQESRLCSIALSGMLG 116
 143 SAISFNHPPTDIFVNVQAVGALGVLYPANHGWMKFWAPLERKLAFTFTGSSVG 202
 117 CFTAILIGFISETLGMPFVYIFGVCVCLLMFVYIDDPVSPMISTSEKEYIIS 176
 203 VMGLASATLVSHFSKTFYFGVYGIIMSLIMVYSSHPETHOYISDDEKKOY--- 260
 177 LKQVGS-----SKOPLPIKMLRSLPIWSICLGFHQMLVSTWVYIPTYISSYVHV 230
 260 -TEKIDVAVKNNSLTTLPRDMWTSASVAWAIITCFECSRMGFELLGQNLQVLMKVDLHI 318
 231 NIDNLLSLAPFIVAMVGMVGLADELLTR-KRLTIYKRIATILGSLPSALIVSL 289
 319 DINSGPISIFPQMGCIYVATGQLDYLRSSGKMSSTEAVKSVNTFGFTVEMAGCL 378
 290 PYLNSGYTATALLTSCGISTLCOSGIYINVDIAPRYSFSLGASRGFSIAPV---I 346
 379 AFPRDVIVANTCLV-IACIGSSSVLSGFNNHDIAPRIVAPILMGIANGIGAAVAGVGMV 437
 347 VPTSGFLLSQDPEFGMRNVFLLPAVNILGLLFYILFEGADVQEMAKE 395
 438 TMTVT-----YQNDP-GMKWVFLAMADIRGVIFLIFAKGDVLPVARE 481

RESULT 10
 YKH4.CAEEL
 ID YKH4.CAEEL STANDARD; PRT; 568 AA.
 AC P34272;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DN HYPOTHETICAL 63.0 KD PROTEIN CO2C2.4 IN CHROMOSOME III.
 GE CO2C2.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdittia; Rhabdittidae;
 OC Rhabdittina; Rhabdittidae; Rhabdittidae; Pelodertinae; Caenorhabdittis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Crakon M., Dear S., Du Z., Durdin R., Favey L., Fraser A.,
 Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 Johnston L., Jones M., Kershaw J., Kirsten J., Laisler N.,
 Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Ritken L., Roopa A., Saunders D., Shonkhen R.,
 Sims M., Smalton N., Smith A., Smith M., Sonhammer E., Staden R.,
 Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
 Watson R., Watson A., Wainstock L., Wilkinson-Sprout J.,
 Wohldman P.;
 *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans.;
 RL Nature 368:32-38(1994).
 CC -I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -I- SIMILARITY: TO MAMMALIAN SODIUM/PHOSPHATE CO-TRANSPORTER 1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL: L23649; AAA37912.1; -
 DR PIR: S44742; S44742; CE00030.
 DR WORMPEP: CO2C2.4; CE00030.
 KW Hypothetical protein; Transport; Transmembrane; Glycoprotein; Symport;
 KM Sodium transport.
 FT TRANSMEM 50 70 POTENTIAL.
 FT TRANSMEM 140 160 POTENTIAL.
 FT TRANSMEM 177 197 POTENTIAL.
 FT TRANSMEM 200 220 POTENTIAL.
 FT TRANSMEM 232 252 POTENTIAL.
 FT TRANSMEM 258 278 POTENTIAL.
 FT TRANSMEM 325 345 POTENTIAL.
 FT TRANSMEM 364 384 POTENTIAL.
 FT TRANSMEM 401 421 POTENTIAL.
 FT TRANSMEM 428 448 POTENTIAL.
 FT TRANSMEM 452 472 POTENTIAL.
 FT TRANSMEM 496 516 POTENTIAL.
 SQ SEQUENCE 568 AA; 63038 MW; 5A3C65642D22F11F CRC64;

Query Match 13.5%; Score 280; DB 1; Length 568;
 Best Local Similarity 22.8%; Pred. No. 4.8e-14;
 Matches 92; Conservative 68; Mismatches 191; Indels 52; Gaps 10;

5 ETLIPKRVPSLCARVGIALLVLF-----HCFNTTIANQYINIMITAMVNSTSPQS 58
 161 ERSPRHVLDISVALYLVVTPPLATVHFGYFVFLARIGM----- 203
 59 DSSEVLVPVDFSGISKAPKSLPAKSSILGQFAIWERMGPPQESRLCSI-ALSGMLG 117
 203 -----GLGEG-FYFPTNNALIG-----NMFPSSSEKSTALSTFTLGQNLASA 242
 118 FTAIIIGFISETLGMPFVYIFGVCVCLLMFVYIDDPVSPYPISTSEKEYIISL 177
 243 AGSPMVAAVACASDLGWPATFYFAGIFATGMSILMFFPAASHPAKVMKTKKEKEYLLANV 302
 178 KOVGSCK--QPLPIKMLRSLPIWSICLGC-FSHQMLVSTWVYIPTYISSYVHVNRD 234
 303 VKRVHSEKTRSIPIYSKILTS-PAFLQLOCHPEFVNLFMTLFOIYIPLSYKEVYLHLGVIA 361
 235 NGLLSALPFIVAMVGMVGLADELLTRKRLIT-VKRIATILGSLPSALIVSLPYLN 293
 362 NGFTALPNTFNMIFKVVWGIGIDKLENKILNSTKAVKSHGVASGSSFSCLLAFEV 421
 294 SGYTTATALLTSCGLSTL--COSGIYINVDIAPRYSFSLGASRGFSIAPVIPTVS 351
 422 DCSNPTTGLIFCLMYSMGTFVSGFTSLSLAPQYATATMSAISMVAMIGRTTPAVM 481
 352 GFLISQDPEFGMRNVFLLPAVNILGLLFYILFEGADVQEMAK 394
 482 SMFRKDGTAEMONIFGCSIAHIFSGSIFLFTSGELQDMAK 524

RESULT 11
 ID GUDT_ECOLI STANDARD; PRT; 450 AA.
 AC Q46916;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)

QY	299	ATALLTECGSTLCOQSSTIYINVL-----DIAPYSSFLMAGSNG-----FSSIPAVIP	348
Db	346	NTTLVVM---LMALAFBFGKGCAGLGMPIYSIDPAEE---IYGLCGGVNFEQSNASIYTP	399
QY	349	TVSGFLSODEPFCGRNVFFLLFAVNLGLLEFYI-IFGEADVOEMAK	394
Db	400	LVIGIYVSELHSFNALVFVGCSSA--LMAWVCYLFVVGDIKRMELQK	444
RESULT	13		
GUPT_BACSU	STANDARD:	PRT:	455 AA.
AC	P42237:		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	01-NOV-1995 (Rel. 32, Last annotation update)		
DE	PROBABLE GLUCARATE TRANSPORTER.		
GN	YCBE.		
OS	Bacillus subtilis.		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;		
CC	Bacillus/Staphylococcus group; Bacillus.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-168;		
RX	MEDLINE: 95219079.		
RA	Ogawa K., Akagawa E., Nakamura K., Yamane K.;		
RT	"Determination of a 21548 bp nucleotide sequence around the 24		
RL	degrees region of the Bacillus subtilis chromosome."		
CC	Microbiology 141:269-275(1995).		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
CC	-1- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/)		
CC	or send an email to license@isb-sib.ch).		
DR	EMBL: D30808; BAA06469.1; -		
DR	EMBL: 299105; CAB12042.1; -		
DR	SUBTILIST: BGL1160; YCBE.		
DR	PRFM: PF00083; sugar_tr. 1.		
KW	Hypothetical protein; Transmembrane; Transport.		
KM			
FT	TRANSMEM 19 39	POTENTIAL.	
FT	TRANSMEM 59 79	POTENTIAL.	
FT	TRANSMEM 87 107	POTENTIAL.	
FT	TRANSMEM 108 128	POTENTIAL.	
FT	TRANSMEM 153 173	POTENTIAL.	
FT	TRANSMEM 177 197	POTENTIAL.	
FT	TRANSMEM 253 273	POTENTIAL.	
FT	TRANSMEM 289 309	POTENTIAL.	
FT	TRANSMEM 320 340	POTENTIAL.	
FT	TRANSMEM 348 368	POTENTIAL.	
FT	TRANSMEM 386 406	POTENTIAL.	
FT	TRANSMEM 414 434	POTENTIAL.	
SO	SEQUENCE 455 AA; 49254 MW; ADDFD3F569CC7B8 CRC64;		
Query Match	12.2%;	Score 252.5;	DB 1; Length 455;
Best Local Similarity	23.8%;	Pred. No. 4,4e-12;	
Matches	94;	Conservative 74;	Mismatches 154; Indels 73; Gaps 16;
QY	19	RYG---TALVLFHCNFTTIAQN-----VTMNTVMVAVNSTSPQSLNDSSEVL	64
Db	81	RFGSKTIIALSIFFWSFTLLQGAIGFSAGTAIILLFLRLFY-----	125
QY	65	PVDKSGIKAPKPSLPANSSIIIGGFAIMERGPQOERSRLCSIALSGMLIGCTAILIG	124
Db	125	-----GISEAP-SFGNGRVVAS-----WFPSSERTASAFNSAOFYAVIIPPLM	170

Oy	125	GEISSTLPMPEFYIFGCGVCCLLMPFVIYDDPVSPWISTSEKRY-----IS---	176
Dd	171	GMLTSHPCGHSHFFVVMGIAGILLIAVIMLKTY-PEKKHPKNVEALAIIEGGGLISDD	229
Oy	176	SLKQGVSSSKOPLPPIKAMLRSLPWSICIGCFSSHOMLVSTM---VVYIPYYISSVHYN	231
Dd	230	SXSGKETESKMPPY-IKOLLTN---RMILGYIAQYCTTLTYFLFMFPYVLVGARMS	284
Oy	232	IRDNCLLSAPETIVAMVIGMNGGYIADPLTRKKFPLTVRKTIATITLSLPSALIVSLPY	291
Dd	285	IIEAEFVASLPALPCGFAGVIGLGIVSDILKLKGRSLTFEARKPPIIANGMLSCSMIV-CNY	343
Oy	292	LNSGITFTF--ALLTLISGLSTLCOSGIVINLDIAPRYSSTLMGA-SRGFSSTAPVTP	348
Dd	344	TDSANLVVYVIMSLEAFPGKGFGALG---WAVSPTSPECAGLSGLFNTEGNIASITTP	399
Oy	349	TVSGFLLSODEPFEGMRNVFEFLFAVNLLGLELYLI 383	
Dd	400	IIGYIVNATGSFNGALVF--VGANAIAAIIISYLL 432	
RESULT	14		
ID	DGOT_ECOLI	STANDARD:	PRT; 445 AA.
AC	P31457:		
Dt	01-JUL-1993	(Rel. 26, Created)	
Dt	01-JUL-1993	(Rel. 26, Last sequence update)	
Dt	15-JUL-1998	(Rel. 36, Last annotation update)	
DE	D-GALACTONATE	TRANSPORTER.	
GN	DGOT.		
OC	Escherichia coli.		
OC	Bacteria; Proteobacteria;	gamma subdivision; Enterobacteriaceae;	
OC	Escherichia.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-K12 / MG1655;		
RX	MEDLINE; 93315143.		
RA	Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;		
RT	"DNA sequence and analysis of 136 kilobases of the Escherichia coli		
RL	genome: organizational symmetry around the origin of replication.";		
RL	Genomics 16:551-561(1993).		
RN	[2]		
RP	SIMILARITY TO DAL5 FAMILY.		
RA	Koonin E.V.;		
RL	Unpublished observations (OCT-1993).		
CC	-I- FUNCTION: INTAKE OF GALACTONATE INTO THE CELL.		
CC	-I- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE		
CC	(PROBABLE).		
CC	-I- SIMILARITY: BELONGS TO THE PHTHALATE PERMEASE FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; L10328; AAA62043.1; --		
DR	ECODGENE; EG11715; DGOT.		
KM	PFAM; PF00083; sugar_tr; 1.		
DR	Transmembrane; Inner membrane; Transport.		
FT	TRANSMEM	33	53
FT	TRANSMEM	70	90
FT	TRANSMEM	100	120
FT	TRANSMEM	158	178
FT	TRANSMEM	183	203
FT	TRANSMEM	258	278
FT	TRANSMEM	299	319
FT	TRANSMEM	325	345
FT	TRANSMEM	353	373
FT	TRANSMEM	391	411
FT	POTENTIAL.		

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 25, 2000, 14:11:04 ; Search time 50.3 Seconds
(without alignments)
552.743 Million cell updates/sec

Title: US-09-391-958-1
Perfect score: 2076
Sequence: 1 MOVDELIRPKVPSLCSARY.....LIRGADVQEMAKERKLTRL 401

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :
1: SP_archaea:*
2: SP_bacteria:*
3: SP_fungi:*
4: SP_human:*
5: SP_invertebrate:*
6: SP_mammal:*
7: SP_mhc:*
8: SP_organelle:*
9: SP_phage:*
10: SP_plant:*
11: SP_rodent:*
12: SP_virus:*
13: SP_vertebrate:*
14: SP_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1023.5	49.3	497	4	O9Y2C5 homo sapien
2	485.5	23.4	483	5	O61369 drosophila
3	475	22.9	560	11	O62634 rattus norv
4	449	21.6	573	5	O09932 caenorhabdi
5	436	21.0	512	10	O82330
6	404.5	19.5	524	5	O94886
7	388	18.7	413	10	O23065
8	363.5	17.5	501	5	O16923
9	320	15.4	516	5	O10048
10	306.5	14.8	499	5	O01791
11	304.5	14.7	530	5	O01537
12	282.5	13.6	506	5	O23576
13	279.5	13.5	420	5	O23558
14	279.5	13.5	592	5	O93599
15	259	12.5	478	5	O23063
16	254.5	12.3	543	5	O44130
17	247.5	11.9	516	5	O94XCL
18	247	11.9	537	5	O17785
19	244	11.8	455	5	O76838
20	244	11.8	473	5	O16252

21	243	11.7	493	5	O09479	O09479 caenorhabdi
22	227.5	11.0	462	5	O76837	O76837 caenorhabdi
23	219.5	10.6	419	5	O9XVEO	O9XVEO caenorhabdi
24	218	10.5	479	5	O23012	O23012 caenorhabdi
25	211.5	10.2	485	5	O9XVKO	O9XVKO caenorhabdi
26	210	10.1	345	2	P94774	P94774 erwinia chr
27	206.5	9.9	493	5	O22832	O22832 caenorhabdi
28	205.5	9.9	421	5	O9XUV7	O9XUV7 caenorhabdi
29	203.5	9.8	464	5	O09484	O09484 caenorhabdi
30	200.5	9.7	466	5	O16976	O16976 caenorhabdi
31	198	9.5	552	5	O21538	O21538 caenorhabdi
32	192.5	9.3	406	5	O94307	O94307 caenorhabdi
33	190.5	9.2	422	2	O34456	O34456 bacillus su
34	185	8.9	467	5	O20265	O20265 caenorhabdi
35	183	8.8	452	5	O61803	O61803 caenorhabdi
36	179	8.6	498	5	O21312	O21312 caenorhabdi
37	173	8.3	332	2	O92FRL	O92FRL burkholderi
38	172.5	8.3	452	5	O44595	O44595 caenorhabdi
39	172.5	8.3	482	5	O94305	O94305 caenorhabdi
40	164.5	7.9	501	11	O9W081	O9W081 mus musculu
41	164.5	7.9	583	5	O18935	O18935 caenorhabdi
42	160.5	7.7	599	5	O22089	O22089 caenorhabdi
43	159	7.7	445	5	O45678	O45678 caenorhabdi
44	157	7.6	596	5	O18472	O18472 caenorhabdi
45	154	7.4	479	5	P90832	P90832 caenorhabdi

ALIGNMENTS

RESULT 1
O9Y2C5 PRELIMINARY; PRT: 497 AA.
AC O9Y2C5;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-NOV-1999 (TREMBlrel. 12, Last annotation update)
DE NA/P04 CONTRASPORTER HOMOLOG.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RX SEQUENCE FROM N.A.
MEDLINE: 99253143.
RA SHITUI A., TSUNODA T., SEKI N., SUZUKI Y., SUGANE K., SUGANO S.;
RT "Isolation and chromosomal mapping of a novel human gene showing
homology to Na+/P04 cotransporter.";
RL J. Hum. Genet. 44:190-192(1999).
DR EMBL: AB020527; BAA76663.1; -
SQ SEQUENCE 497 AA; 54055 MW; 128F47B4 CRC32;

Query Match	49.3%	Score 1023.5;	DB 4;	Length 497;
Best Local Similarity	43.4%	Pred. No. 26-65;		
Matches 202;	Conservative 70;	Mismatches 114;	Indels 79;	Gaps 2;
OY 16	CSARRIGIALVHFCNFTTIAQNVINITYAMVNSTSPQSLDSEVLVDVDFGLSKRA 75			
DB 33	CSVRHRLALILDLCNFSITYOQNLISALPAMVNNTPAPSQPASTERSTDSQGYMNET 92			
OY 76	PSLSPAKS-----			84
DB 93	LKEFRKMAPAYDKSPETIGILLSLWGSFLAPISPGYVAGIRGAKYVVGAGLFISSFLT 152			
OY 84	-----SLGGQFAIMERNGPQERSRLCSIALSGMLGC 117			
DB 153	LFIPLAANAGVALLIVLRIVQIGIAQVAVLTGQYSIVWKNAAPLERSQLTTIAGSGMLGS 212			
OY 118	FRAILIGGFISETLGPFPFYITFGVGCVCCLLMFVYIYDDPSTYPMISTSEKYEIISLS 177			
DB 213	FYVLVAGGLLCQRTIGMPYFYIFGIGCACCPPLMPLITDDPNVHPFISGEXRYIVCSL 272			
OY 178	KQGVSSKQPLPIKAMRLSLPIWSICIGCFSHQWLSTVWVYIPTVISSVYHVNINDGL 237			

Dd	273	AQODCSPEGMSLPIRAMKSLPLMALLVSYFCFYWLFYTIMATPTIYSVLOANLRDSCI	332
Oy	238	LSALPFIYAMVIGMNGYADLPFLTRK-FRLITVRKIATILGLSDPSALLVSPLYNSGY	296
Dd	333	LSALPFVVGCIIILIGGLADFLFRLKILIRLTIRKLIFPAIVLPPSVIVSLPWRRSH	392
Oy	297	ITATALLTSCGLTSCSGIYINVLDAIPRSSFMGMSRGSFSAIPIYTVSGFILS	356
Dd	393	SMTMFVLVSSAISFCESGALVNFLDLAPRTTGFKLLQVFAMHAGASIPFAAGFFIS	452
Oy	357	QDPFGMRNVFELFAVNLGLLFYLFIGEADVOMAKERKLTRL	401
Dd	453	QDSFGMRNVFELLSAAVINISGLVFLIFGRADYOWAKROTFTHL	497
RESULT	2		
ID	061369	PRELIMINARY;	PRT; 483 AA.
AC	061369;		
Dt	01-AUG-1998	(TREMBLrel. 07, Created)	
Dt	01-AUG-1998	(TREMBLrel. 07, Last sequence update)	
Dt	01-NOV-1999	(TREMBLrel. 12, last annotation update)	
DE	PUTATIVE INORGANIC PHOSPHATE COTRANSPORTER.		
GN	PICOT.		
OS	Drosophila anaassae (Fruit fly).		
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
CC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;		
OC	Ephyroididea; Drosophilidae; Drosophila.		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-TAI 13-1610;		
RA	DA LAGE J.-L., ALLAND C.;		
RL	Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.		
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.		
DR	EMBL; AF024691; AAC39088.1; -		
DR	FLYBASE; FBgn0024472; Dana\Picot.		
Kw	Transport; Transmembrane.		
FT	TRANSMEM 187 207 POTENTIAL.		
FT	TRANSMEM 292 312 POTENTIAL.		
FT	TRANSMEM 383 403 POTENTIAL.		
FT	TRANSMEM 420 440 POTENTIAL.		
SO	SEQUENCE 483 AA; 52885 MW; EBEB3B7C CRC32;		
Query Match	23.4%; Score 485.5; DB 5; Length 483;		
Best Local Similarity	27.5%; Pred. Max 3.3e-27;		
Matches 123; Conservative	81; Mismatches 163; Indels 81; Gaps 11.		
Oy	24	LVLHPCNETTIQAONIYNITVAAYNST--SPQSOLDN--SSEVLPVDS-----	69
Dd	2	LFLGMANNVYMNTN--MSAIVAAMNHNAIKGEDEYEDDECDDROIPIDDSGDERPMWA	59
Oy	69	-----EGLSKAPKSLP-AKSSIIGQFA-----IW--	94
Dd	60	ALOGYILSSFEGYVITOIPFGILLAKKKGSLFLGYGLMINSVFVAFIVAREGVWGL	119
Oy	94	-----ERMGPQERSRCLSIALSMLLCFTAILLGFISE--	130
Dd	120	CAVRFIQGLEGPVPCTHAMLAKEIIPERSRMGAAYAGAOFSTIISMPLSLAEYG	179
Oy	130	TLAGPFEYVINGVCVCCLMFVIYIDDPSPYMWISTSEKEYIITSLKQOVGSSKOPL	188
Dd	180	FDCGMPSPFYVGIVGTWSIAFLFVEDPSTHRKIDEREKKYINESIMGDIYAKSPI	239
Oy	189	PIKAMLRSLPTSICLGFSSHQMVLSTNYVYIPTIYSYYHNINDGSLAPITVAV	248
Dd	240	PKFSIVKSILPFAIILFAHGHNHYGETETELTPYMKOYLRFSLKSNGLISSLPYLAWL	299
Oy	249	IGMVAGYADLFL-LTKKFLIVRKIATILIGLSPSALLVSLPYNSGYITATALLTSC	307
Dd	300	LSMFIISVIAADMWISSKRSLTRKILINSIGGYGGLALTAASYTCDBRALTLALTIGV	359

```

QY 308 GASTLQSGIYINVLDIAPRYSSEFLMGASRGSSIAPIVPIVSGFELS--GDPERG-WR 364
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 360 GLNGIGYSEFKINHLDPFRFAGFLMSTINCASNLGLAPLAAGNLIDSPSKPYMGQWO 419
QY 365 NVEFLLEFAVNLGLLFYLLFEGADVQEW 392
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 420 IVEFLIAFPIYLLICGTFYNNFGSGEQFW 447

RESULT 3
ID 062634 PRELIMINARY; PRT; 560 AA.
AC 062634:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
DE BRAIN SPECIFIC NA+-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 94261635.
RA NI B., ROSENECK P.R., NADI N.S., PAUL S.M.;
RT "Cloning and expression of a cDNA encoding a brain-specific Na(+)-
RT dependent inorganic phosphate cotransporter."
RL Proc. Natl. Acad. Sci. U.S.A. 91:5607-5611(1994).
DR EMBL; U07609; AAA19646.1; -.
SQ SEQUENCE 560 AA; 61665 MW; C7213CFL CRC32;

Query Match 22.9%; Score 475; DB 11; Length 560;
Best Local Similarity 28.0%; Pred. No. 2.1e-26;
Matches 111; Conservative 66; Mismatches 158; Indels 62; Gaps 6;

QY 12 VPSLCASARGIALVLFHCFETTLAQNVMINIMVAMVNSTSPQSOLDNSEVLPVDSFG 71
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 161 IPSAARVHIGCVI-----FVRILOGVEGVTPAC----- 191
QY 72 LSKAPRSLEPAKSSILGQFAIWERMGPPQERSLCSIALSGMLGCFYALLIGFISETL 131
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 191 -----HGWSKNAPLERSRRLATTACGSYAGAVYAMPPLAGLVQYS 232
QY 132 GMPFVFIYIGGVCVCCLLMFVYIDDPYSYWIISTSEKEYIISLKQOVSS----- 185
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 GMSVYFVYVYSGFIWYLFMLVSYESPALHPSISEEKRYI-----EDAIGESAKLMNPV 288
QY 185 -KQPLPIKMLRLSPIMWSCIGCFSHQWLVSMVYIPIYISSYVHVNIRDGLLSALPF 243
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 289 TEFNTPWRREFTSMYPYALIVANFCRSWTFYLLLSIPAYEEVGVGFELSKVGLVSALPH 348
QY 244 IYAWYIGMWGIAFLADFLILTKKFRLLT-VKIIATILIGSLPSSALIVSLPYLNGYITATL 302
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 349 LVMTIIVIPGGIADFLRSRHMSTTNVAKLMNCGFGMEATLLLVGYSHSKGV-AISF 407
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 303 LTLSCGLSTLCOSGIYINVLDIAPRYSSEFLMGASRGSSIAPIVPIVSGFELSDPERG 362
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 408 LVLAVGSEGFALSGFNVNLDIAPRASIILMISNGVGLSLGMVCPILYIGAMTKTKTRE 467
QY 363 WRNVEFLLEFAVNLGLLFILNGEADVQEWAKERKLT 399
||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 468 WQYFELIASLVHGVGVIYFGVFASGEKQFWAPDEEWS 504

RESULT 4
ID 009932 PRELIMINARY; PRT; 573 AA.
AC 009932:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE HYPOTHETICAL 63.6 KD PROTEIN K10G9.1 IN CHROMOSOME III.
GN K10G9.1.

```

OS Caenorhabditis elegans.
 OC Eukaryota: Metazoa; Nematoda; Secernentea; Rhabdita; Rhabditida;
 OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA MORITZMORE B., BUCK D.;
 RL Submitted (FEB-1995) to the EMBL/Genbank/DBJ databases.
 DR EMBL: 236282; CA85289.1; JOINED.
 DR EMBL: 248055; CA85289.1; JOINED.
 DR EMBL: 248055; CA88135.1; JOINED.
 DR EMBL: 236282; CA88135.1; JOINED.
 SQ SEQUENCE 573 AA; 63610 MW; F00B8CC CRC32;

Query Match 21.6%; Score 449; DB 5; Length 573;
 Best Local Similarity 26.7%; Pred. No. 1.5e-24;
 Matches 125; Conservative 84; Mismatches 165; Indels 94; Gaps 14;

OY 1 MOVDTLIPRKVPISCSAAYGIALVLFH-----CNFTIAOVININIMVAVNST 51
 DB 35 LQEDKKFPLRKV-----RMQIALAHGFAISFGIRSNF-GVAKRMKN-----NFT 80
 OY 52 SPQSLND-----SSEVLPRVDS-FGGLSK-----APKSLPAKSSILGQFA---- 92
 DB 81 DAYGEVHEKEFWTGTGEVGMMESSFFGYAASQIPAGVIAAKFAPKLMGLIFASLLN 140
 OY 92 -----IWERNGPOERSRLCSIALSGML 114
 DB 141 IYTAICLNHPFTDIFVAVIQVMGGLACVCPAMHGVMKWAYPAPLERSKLATTFPTGAS 200
 OY 115 LGGCFALLIGFISLSELPFVEYFEGGVCYCCILMFVYIYDDPSYMWISSEKEXII 174
 DB 201 VGVMVGLPASALVLFHSFSTPFPVFGALGYWSILMTYSGTSPETHGYISADEKKYI- 260
 OY 175 SSLKQOVGS-----SKOPLPIKAMLRSLPIWSICLGFCSHOWLVSIMVYIPTYSYV 228
 DB 260 ---TEKVGSVANKNMTLTLLPWRDMWTSRAVVAIIICSCRSRSPFLILGNOLTYKDWL 316
 OY 229 HVNINDNGLSLAPFIVAVIGMVGQYADPLTK-KRPLITVRKIATILGLSPSSALIV 287
 DB 317 HIDIKNSGLIAIFPOLGMCIVTLTSGQLSDYLRSSGKMSTEARVKNFEGFVEAVMLG 376
 OY 288 SLPIYNSGITTATLLTSCGLSTLQSGIYINVDIARYSSEFLMGARSSSTAPVIV 347
 DB 377 CLAFPRDPIVATFPLI-TACSGAGAVLSCFNVNHDIARHAPILMLINGALAGV-G 434
 OY 348 PTVSGFLLSQDEPFGRNVFLLFAVNLGLFYLLFEGADVOENAKE 395
 DB 435 GIVTNSLTRYQND-GMOWVFLMLMSIDIRGIIFLLFAKGDVLPAPARE 481

RESULT 5
 AC 082390 PRELIMINARY; PRT: 512 AA.
 ID 082390;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, last sequence update)
 DT 01-MAY-1999 (TREMBLrel. 10, last annotation update)
 DE T27A16.25 PROTEIN.
 GN T27A16.25
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
 OC core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
 OC Arabidopsids.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA;
 RA ROUSLEY S.D., LIN X., KAUL S., SHEA T.P., FUJII C.Y., MASON T.M.,
 RA SHEN M., RONNING C.M., FRASER C.M., SOMERVILLE C.R., VENTER J.C.;
 RT "Arabidopsis thaliana chromosome II BAC T27A16 genomic sequence."
 Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

DR EMBL: AC005496; AAC35230.1; -.
 DR PFAM: PF00083; sugar tr. 1.
 SQ SEQUENCE 512 AA; 56498 MW; 933A5143 CRC32;

Query Match 21.0%; Score 436; DB 10; Length 512;
 Best Local Similarity 30.5%; Pred. No. 1.1e-23;
 Matches 101; Conservative 66; Mismatches 140; Indels 21; Gaps 7;

OY 63 VLPRVDSFGLSKAPKSLPAKSSILGQFAIWERNGPOERSRLCSIALSGMLGCTAIL 122
 DB 197 LTVRAFMCVGRG-VAMPAMNNTL-----SKWVPYQERSRLALVSGMYIGVTGLA 248
 OY 123 IGGFISLSELPFVEYFEGGVCYCCILMFVYIYDDPSYMWISSEKEXIISSLKQV 182
 DB 249 FSPFLIHQFGWVSFSGTGLVTLMLTYAESPLDPLDPLPERKLIADNC----- 304
 OY 183 SSKOP---LPKAMLRSLPIWSICLGFCSHOWLVSIMVYIPTYSYVHNINDNGLS 239
 DB 304 ASKEPKSIPKRLSLPPVMAISCHFNWCTFTLLTMPTYYHQLKFNLMESGLLS 363
 OY 240 ALPEIVAVIYGVGYLADPLTKRPLITVRKIATILGLSPSSALIVSLPIYNSGYTA 299
 DB 364 VFPWMTMAISANAGMIADTLVSRGFSVTVNRKIMQITGFLGPAFLTOLKHIDS--PT 420
 OY 300 TALLTISC---GLSTLQSGIYINVDIARYSSEFLMGARSSSTAPVIVPTVSGFLSQ 357
 DB 421 MAVLCACSGQGDPAFSQSGIYNSHODIAPRYSGLVGLSTPAGVLAGVLTATGHIH 480
 OY 358 DPEFGNRNVFLLFAVNLGLFYLLFEGAD 388
 DB 481 G---SMDDVFTISVGLVGYVIMNLFSTGE 508

RESULT 6
 AC 094886 PRELIMINARY; PRT: 524 AA.
 ID 094886;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, last sequence update)
 DT 01-AUG-1998 (TREMBLrel. 07, last annotation update)
 DE NA(+)-DEPENDENT INORGANIC PHOSPHATE COTRANSPORTER.
 GN NAPI-T.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-OVERLAP;
 RA MACIVER B.;
 RA Thesis (1997), University of Edinburgh, Edinburgh, UK.
 DR EMBL: Y07720; CA68992.1; -.
 DR FLYBASE: FBgn0016684; NAPI-T.
 SQ SEQUENCE 524 AA; 58814 MW; 3F888436 CRC32;

Query Match 19.5%; Score 404.5; DB 5; Length 524;
 Best Local Similarity 31.6%; Pred. No. 2e-21;
 Matches 95; Conservative 60; Mismatches 141; Indels 5; Gaps 5;

OY 95 RMGPPOERSRLCSIALSGMLGCTAILIGFISLSELPFVEYFEGGVCYCCILMFVY 154
 DB 170 KWIIPERKFNVS-AVLGSSGVALLPYIFGYIIDTTRMWMVYVIGIYGTLMFIMQFL 228
 OY 155 IYDDPSYPMIISSEKEXIISLQOVGSSKOPLPKAMLRSLPIWSICLGFCSHOWLVS 214
 DB 229 VEPSPAEHRPAIDSEKRFLEKSLGASIGSGKTPPKAITSRPWLNVYAQMGITGLP 268
 OY 215 TAVVYIPTYISSYVHVNINDNGLSLAPFIVAVIYGVGYLADPLTKRPLITVRKI 273
 DB 289 TLTHTAPTYFRLIHHNINIRATGSLGPHLMRLAFVVSIFADVILRTDKMSRTVNRKL 348

OY		274	ATTTGSLPSSALIVLPTLNISGYIATAFLTSCGLSTLCOSGIYINVDIAIPRYSFLLM	333
		: :	: :	: :
Dd		349	ATFTCCCKRGKGLVLTMLAFNGNATPAIYLVIYATMLHGAVSSEPLASMDLSFNAYGIYL	408
		: :	: :	: :
OY		334	GASRCGFSSIIAPYIVPTVSGFLLSDPPEF-GWNRVFPELLFAVNLLGL-LFYLIIFGEADVOE	391
		: :	: :	: :
Dd		409	GVSQMGIGMPGRFISPFIYGQLTHNNQTIDAMKNV-ELLTLTMLTGSGILYYVFSKSLOP	467
		: :	: :	: :
OY		392	W 392	
Dd		468	W 468	
RESULT	7			
023065				
ID	023065	PRELIMINARY;	PRT,	413 AA.
AC	023065;			
DT	01-JAN-1998 (TREMBREL. 05, Created)			
DT	01-JAN-1998 (TREMBREL. 05, Last sequence update)			
DT	01-AUG-1998 (TREMBREL. 07, Last annotation update)			
DE	BAC IG005710.			
GN	A-IG005110.NN.			
OC	Arabidopsis thaliana (Mouse-ear cress).			
OC	Euraryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	eunhylliphytes; Spermatophyta; Magnoliophyta; eudicotyledons;			
OC	core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;			
OC	Arabidopsids.			
RN	[1]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-CV. COLUMBIA;			
RC	ANDREWS S.;			
RA	Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN-CV. COLUMBIA;			
RC	WATERSTON R.;			
RA	Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.			
RL	EMBL; AF013293; AAB62846.1, -.			
DR	SEQUENCE 413 AA; 45791 MW; OD120117 CRC32;			

	Query March	18.7%	Score 388;	DB 10:	Length 413;
	Best Local Similarity	25.7%;	Pred. No. 2.4e-20;		
	Matches 106;	Conservative	69;	Mismatches 133;	Indels 104;
					Gaps 11.
QY	19 RYGIALVLFHCFNFTTIAQNVININIMVAVNVSSTPSQO-LNDSSEVLPV-----	67			
DB	60 RRMVIVLVCFSF-----LLCNMDRVNMSIALIPMSQGYNMSSATVGLIGSFFWGYLL	113			
QY	67 -----DSGGGLSKAPKSLPAASSILIGGFALMERKGPQERSRLCSIALSGMLGC	117			
DB	114 TQILGGIMADKFGGKGA--MPANNMNL-----SKMIPVSEKSRSLATVYSGMYTGS	163			
QY	118 FTAILIGFISFTLMEPEFYIFEGVGCCLLMEVYVIDDPVSYPMISTSEKEYITSL	177			
DB	164 VTGLAFSPMLIRKPMPSVTFYFSGSLGSIWFLMLLKAFASSPKDDPDLSEKKYILGSS	223			
QY	178 KOQVSSKOP--LPKAMLRSLPIWISICLGFCSHOWLVSTVWYIPTY-----T	224			
DB	224 K-----PREPVVIMPKLLISKRPVWALLIISHFCNMWGTFILLTMPTPYNOARSSASYI	278			
QY	225 SSVYHV-----NIRDNGILSALPEIYAWYIGWVGVLADFLITLKKRLITVKRIATI	276			
DB	279 SLTFNIFCEQVILKFNITESGLICVLPWLMAPANIGMIATLTVLSRG-----	327			
QY	277 LGSIPSSALIVSLPYLNSGYIRATALLTILSCGLSTLCGSIYINVDIAPRYSSFLMGAS	336			
DB	327 -----LSTINSDAFAFSQSGELVSNHQDIGPRAGVLLGSS	360			
QY	337 KGFSSIAPIYIVTVSGFLLSOPEFGWRVFLFLRAVNLGILFYILFEAD	388			
DB	361 NTAGVLAGVGTAAATGYILQRG---SMDVYFEVNAVALYILGLVWMLFTGCE	409			

RESULT	8
ID	016923
AC	016923; PRELIMINARY;
DT	01-JAN-1998 (TReMBLrel_05, Created)
DT	01-JAN-1998 (TReMBLrel_05, Last sequence update)
DT	01-NOV-1998 (TReMBLrel_08, last annotation update)
DE	F25G6.7 PROTEIN.
GN	Caenorhabditis elegans.
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdittia; Rhabdittidae;
OC	Rhabdittina; Rhabditoidea; Rhabdittidae; Peloderinae; Caenorhabditis.
RP	[1]
RP	SEQUENCE FROM N.A.
KX	STRAIN-BRISTOL N2;
KX	MEDLINE; 94150718.
RA	WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA	BONFIELD J., BURTON J., CONNELL M., COSEY T., COOPER J., COULSON A.,
RA	CRAWFORD M., DEAR S., DU Z., DURBIN R., FAVELL A., FULLON L.,
RA	GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIEH M., JOHNSTON L.,
RA	JONES M., KESHAW J., KRISTEN J., LAISTER N., LATREILLE P.,
RA	LIGHTING J., LOYD C., MCCURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA	PASONS J., PERCY C., RIFFEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SILSTON J.,
RA	THERIY-MEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA	WATSON A., WEINSTOCK L., WIKKINSON-SPROAT J., WOHLDMANN P.;
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT	elegans.";
RL	Nature 368:32-38(1994).
RN	[2]
RN	SEQUENCE FROM N.A.
RP	STRAIN-BRISTOL N2;
RC	NELSON J., WOHLDMANN P.;
RA	Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RL	[3]
RN	SEQUENCE FROM N.A.
RP	STRAIN-BRISTOL N2;
RA	WATERSTON R.;
RL	Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
EMBL:	AF022973; AAC25800.1; ..
OR	SEQUENCE 501 AA; 55339 MW; 2D9C2975 CRC32;

Query Match	17.5%	Score 363.5;	DB 5;	Length 501;
Best Local Similarity	23.7%	Pred. No. 1.6e-16;		
Matches 108;	Conservative	98;	Mismatches 165;	Indels 85;
				Gaps 13;
QY	17	SARGIALVLEHCNFTTIAQNVIMNTVWAVNS-----TSPOSQ--LNDSE-----	63	
		: : : : : : : : : : : : : :		
Db	25	SMREFTSITMLCFRCGCVHLMNSNMGMALTCMVNSATYDNETYPPENAPPLDMSDEOGY	84	
		: : : : : : : : : : : : : :		
QY	63	-----VLPDSFGGHS-----KAPKSL-----	80	
		: : : : : : : : : : : :		
Db	85	IFSAPNAGLILWLTFTGGADKFNKATYMLLVSAVLSLNFPLPMAAPISYVAIRSRFLV	144	
		: : : : : : : : : : : : : :		
QY	80	-----PAKSSILGGQFAIWERMPQERSRLCSIALSGMLGCFPAIILI---GCFIS	128	
		: : : : : : : : : : : : : :		
Db	145	GPADALLDPANNSLT-----TRNFPISERSYALGLTFTGGROIG--PLIITIPAGALICS	195	
		: : : : : : : : : : : : : :		
QY	129	ETL---GMPFEVYFIEGGVGCYCCLLMFVITDDPVSYPMISTSEKEYI--ISSLKQVGS-	184	
		: : : : : : : : : : : : : :		
Db	196	QTEITFGGMPISFYLSGFIQVLEIFSYIFILGADKPKSKQCSIDNELKFTTINSQSDYCKK	255	
		: : : : : : : : : : : : : :		
QY	184	-SKOPLPIKAMLRSLPIWSICLGCPSHOMLVSTNVVYIPYISSVYHVNIRDGLLSAP	242	
		: : : : : : : : : : : : : :		
Db	256	RTERKVPKNNKILSGSAVWASVSLVCHEPFLMTLIMFLPSYLIHDVHNHSTENGSLSP	315	
		: : : : : : : : : : : : : :		
QY	243	FIVAVIGMWGGYADLF--LTKKFLRITVRKIATILGSLPSSALISLPYLSNGYI--TAT	300	
		: : : : : : : : : : : : : :		
Db	316	TVSLMFATIGSSYLNTWILQKNKNTWKKDKPICVKNLSIGLGVFLAATFLDNEHAMAV	375	
		: : : : : : : : : : : : : :		
QY	301	ALLTISGLSTLCOSGIYIINVDIAPRYSSFLMGASRGSSSTAPYIVTPVSGELLSDPE	360	
		: : : : : : : : : : : : : :		

Db 376 LFLCLSMASAGLHTPGCOLALVSAPVASCAGVTGTFEFAVSGIHLITKMIYKDRTA 435
Qy 361 FGRNVEFLFAVNLGLFLYLINGEADVOEMAKER 396
Db 436 AEMLVETISTVAIFPIIFNWGSTVOWMAKSK 471

RESULT 9
ID Q10048 PRELIMINARY; PRT; 516 AA.
AC Q10048;
DT 01-JAN-1999 (TREMblrel. 09, Created)
DT 01-JAN-1999 (TREMblrel. 09, Last sequence update)
DT 01-JAN-1999 (TREMblrel. 09, Last annotation update)
DE HYPOTHETICAL 56.0 KD PROTEIN T09B9.2 IN CHROMOSOME X.
GN T09B9.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA KERSHAW J.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: TO SODIUM/PHOSPHATE COTRANSPORTER.
DR EMBL; Z47070; CAAB7340.1; -.
DR WORMPEP; T09B9.2; CE01650.
DR PFAM; PF00083; sugar_tf; 1.
KM Hypothetical protein; Transmembrane.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 304 324 POTENTIAL.
FT TRANSMEM 351 371 POTENTIAL.
FT TRANSMEM 377 397 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 445 465 POTENTIAL.
SQ SEQUENCE 516 AA; 55990 MW; F6DB2754 CRC32;

Query Match 15.4%; Score 320; DB 5; Length 516;
Best Local Similarity 25.4%; Pred. No. 2e-15;
Matches 105; Conservative 78; Mismatches 177; Indels 54; Gaps 13;

Qy 29 CN-----FTIAQNV-----NITMVAVNSTSPQSLND-----SSEVLPV 66
Db 71 CNHKKOFAMSSIOGLIYSGNPFSLFVITGMQADRLNGKMTIYAAMAFITVSNVLP 130
Qy 67 D-----SFG-----GLSKAPKSLPAKSSILGQFAIWMERGPOERSRLCSIALSGML 114
Db 131 SAGASFALVFLRVLTGTGDLALS-PASSSLI-----TMFPKPKRPSALGITSRQ 182
Qy 115 LGCTAIIIGFT-----SETL-GMPFVYIFGGVGCVCCLIMVYIYDDPVSYTWISTS 168
Db 183 IGTLLIPIGIMLCSGSDSKFLGMPALFYLSVVAALVIMVVFSSADKPSKHLICSHN 242
Qy 169 EKEYIISLKQO-VG--SSKQPLPKAMLRSLPMSICLGFSHOMLVSTWVVIPTYSIS 225
Db 243 EEAIVNKRKEENIGKRNKRNTPKALFTSKQVWVAALVCHPEPLVIMLQFLPKPFS 302
Qy 226 SVYHVNIRNDGLSALPEIVAMVIGMGVYLADELTKKF-RLITVRKIATIIIGSLPSA 284
Db 303 DVLGLSTNVNGLVSLPALPIFLSKCLASLASVLTANGYLRKQSCIFNFIASLGIGI 362
Qy 285 LIVSLPLTANS--GYTTAATLTLSCGLSTLCQSGIYINVDIAPRYSSFLMGASRGFSST 342
Db 363 CIAATPLMSNQHAIIWALITIIICLANAFGLTPGVLTAVIOLAPAFSGIITGLAFVASC 422
Qy 343 APVIVPVTSGFLSDDPFGMRNVEFLFAVNLGLFLYLINGEADVOEMAKER 396

Db 423 FSIENKLIISQILRTGSKHEWTIVEISAFVAILPTIEFTLMGASERTEMASNR 476
RESULT 10
ID 001791 PRELIMINARY; PRT; 499 AA.
AC 001791;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-JUL-1997 (TREMblrel. 04, Last sequence update)
DT 01-NOV-1998 (TREMblrel. 08, Last annotation update)
DE SIMILAR TO SODIUM DEPENDENT-PROSPHATE TRANSPORTER.
GN FL2B6.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERS M.,
RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMER E., STADEN K., SULLSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans".
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA PAUIEY A., MAGGI L.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA WATERSTON R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003138; AAB54154.1; -.
DR PFAM; PF00083; sugar_tf; 1.
SQ SEQUENCE 499 AA; 54949 MW; 39AED731 CRC32;

Query Match 14.8%; Score 306.5; DB 5; Length 499;
Best Local Similarity 24.5%; Pred. No. 1.8e-14;
Matches 80; Conservative 72; Mismatches 134; Indels 41; Gaps 7;

Qy 79 LPAKSSILGQFAIWMERGPOERSRLCSIALSGMLCFTAILIGFISETL----- 132
Db 166 IPAGSVLISS-----WPLSEKSTMAIITFTGQIG-----INSMPLTALCOLHFE 214
Qy 132 GMPFVYIFGGVGCVCCLIMVYIYDDPVSYTWISTSSEKEYIISLKQOVGSSKQPLPIK 191
Db 215 GMPLFIYIYGLGAVFLVIMVHRLADKPRESKYITATELTYI-----KGGKQRNRAE 267
Qy 192 AMRLSLP-----INSICLGFSHOMLVSTWVVIPTYSISVYHVNIRNDGLSALP 242
Db 268 TIVRTFPMKIIILNCGVCAICSPASQSFVALVATYLPKYNQIAFKKMLTNGIWSLSLP 327
Qy 243 FIVAVNIRNDGLSALPEIVAMVIGMGVYLADELTKKFRLITVRKIATIIIGSLPSA 302
Db 328 FFIQGITLILFRIIDKVKQRKVNATATKYSNMAIASFSAFIYIAY---GPPDSAE 384
Qy 303 LTLSCGLSTLCQS---GIYINVDIAPRYSSFLMGASRGFSSTAPVIVPVTSGFLSD 358
Db 385 VOLSTIVMAAFSAVVPNTSIVVAPQFTAFISSVQVLAQIASTLAPIVIGRTSHG 444
Qy 359 PFGMRNVEFLFAVNLGLFLYLINGEADVOEMAKER 396

DB 445 TTYEKCAFYSIAGVATGLIFOLF 471

RESULT 11

ID 001537 PRELIMINARY; PRT: 530 AA.

AC 001537

DT 01-JUL-1997 (TREMBlrel. 04, Created)

DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE COSMID F21F8.

GN F21F8.11.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,

RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN T., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULLSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans".

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA WILSON R., FAVELLO A., LE T.T.;

RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U97000; AAC47997.1; -.

DR PFAM; PF00083; sugar_tr; 1.

SQ SEQUENCE 530 AA; 58783 MM; 6C36C5A0 CRC32;

Query Match 14.7%; Score 304.5; DB 5; Length 530;

Best Local Similarity 22.9%; Pred. No. 2,6e-14;

Matches 110; Conservative 69; Mismatches 186; Indels 115; Gaps 15;

DB 15 LCSARIGIALVH-FCNFTIAONVIM-----TTMYAVNVSSTPQSLNDS 60

DB 38 MCOAHAGICLSICMCNSTAYV---LMTNNAITVEGTESTILMSTLEKNTSESQELGCG 94

DB 61 S-----EVLVDVSFGG----- 72

DB 95 DDEGSCOTLKEGVINDYDGTFTMSVSGQYIVSAFLGFIYSPAGVLVDRFSARHILS 154

DB 72 -----LSKAPKSLPAKSLILG--GQFA-----IWEWGPQERSRLC 106

DB 155 VALIMLTJLASLMPVLSITIGEGKAFAGFVNGISETMLIPINSVWTWIPINEXSLA 214

DB 107 STALSG-MLLGCFATILIGFISETLGMPFYIFGVCVCCCLMFVVIYDPPVPMI 165

DB 215 SVFTAGNOLSGFNGILVALCALSSFGWSSIFYSASLFGISMLVLMHLVTRNSPHNTKMI 274

DB 166 STSEKEYIISLKQVGSS-KQPLPKALRLSLPIWSICLGFESHQMLVSTWVYIPTYI 224

DB 275 HRELDIYLNIPKHPKSVVKTWRDMLTSKVEWLSMENSVMGMALALFVYIPYVE 334

DB 225 SSYVHNITDNGLSLPLPTIVAVMGVGLADPLLTTRK-----RLITVKRIATITLSL 280

DB 335 KDVLMDVQNSNGFYSAIPHSIMLAKLWGLYLMKDRKHKIILSPATVRLSOFASMG-- 393

DB 281 PSSALIVSLPYNSGYITR-----ALLTSLC-----GLSTLCGSIYINVLADPRYSFL 332

DB 393 -TSVSCGFILRYMN-----CATPRYALVLLSSVSAFFGLSI-----SGFTYSLISAPSHITGL 444

DB 333 MGASRGFSIAPVIVPVSGLLSODPEFGWRNRPFLFAVNLCLLFLYIFGEADVQEM 392

DB 445 TSLATVIGFVGHMPTPLMISYKYKTIVGTAEMWGHVLLIYFASATGIIIFLFGSDVQNM 504

RESULT 12

ID Q23576 PRELIMINARY; PRT: 506 AA.

AC Q23576

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBlrel. 08, Last annotation update)

DE SIMILAR TO SODIUM/PHOSPHATE TRANSPORTER. NCBI GI: 1086815.

GN ZK682.2.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;

OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RC MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPESEY T., COOPER J., COULSON A.,

RA CRAYTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN T., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,

RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULLSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.,

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans".

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC DU Z., LE T.T.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RC WATERSTON R.;

RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; U41110; AA82415.1; -.

DR PFAM; PF00083; sugar_tr; 1.

SQ SEQUENCE 506 AA; 55385 MM; 8B04E7F9 CRC32;

Query Match 13.6%; Score 282.5; DB 5; Length 506;

Best Local Similarity 24.6%; Pred. No. 9,2e-13;

Matches 82; Conservative 70; Mismatches 140; Indels 41; Gaps 11;

DB 85 ILG-GQFAIM-----ERWGPQERSRLCSIALSGMLGCTAILIGFISETL----- 132

DB 158 VMGFGGVLMPCMLVILVIAQMFNPKSTALATATGNOL-----SVIANFAFAELCOLPW 213

DB 132 GMPFVYIFGVCVCCCLMFVVIYDDPY-SYPTWISTSEKEYIISLKQVGSSKQPLPI 190

DB 214 GMPMFHVAVCGIYWCICLIWMYVVDSPCHADKKLSRDELATITR---ERYLRLRQHPNM 270

DB 191 KAMLRSLPIWSICLGFESHQMLVSTWVYIPTYISVYVNVNIRDNGLSALPLFIAMVYG 250

DB 271 MALMKSPPVWALIAASSFAHNVYVGTITLPLTYKIVLMKSLTNSGLMSALPFLQLLSK 330

DB 251 MYGVLADFLTRKFRITVRKIATITLSLPSALIVSLPYL-----NSGYITATLIT 304

DB 331 VFYAGMAE--SARKRWMSINDITRKCNSASAFGIALICGLCLDCDSORG-----AATF 383

DB 305 LSCGSLTLCOSGIYI-----NVLADIAPRYSFLMGASRGFSIAPVIVPVSGLLSODP 359

DB 384 LIC--LAMCFVSGVIGPYNTSAVTIAPGQTAIAAFAFRFQGIASSVAPYHIGAVTKOCT 441

DB 360 EFGWRNRPFLFAVNLCLLFLYIFGEADVQEM 392

Db 442 ADEWKIVEAVIAIAICVYTGIFOCGTASLDQM 474

RESULT 13

ID 023558 PRELIMINARY; PRT; 420 AA.

AC 023558;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE ELGGANS COSMID ZK652.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

RN Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2; TISSUE-WHOLE ANIMAL;

RX MEDLINE; 92168156.

RA SULSTON J., DU Z., THOMAS K., WILSON R., HILLIER L., STADEN R.,

RA HALLOAN N., GREEN P., THIERRY-MIEG J., QIU L., DEAR S., COULSON A.,

RA CRAXTON M., DUREIN R., BERKS M., METZSTEIN M., HAWKINS T.,

RA AINSCOUGH R., WATERSTON R.;

RT "The C. elegans genome sequencing project: a beginning.";

RL Nature 356:37-41(1992).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2; TISSUE-WHOLE ANIMAL;

RA DU Z.;

RL Submitted (MAY-1993) to the EMBL/GenBank/DBJ databases.

DR EMBL; L14429; AAA28211.1; -

SO SEQUENCE 420 AA; 46281 MW; 4E12C821 CRC32;

Query Match 13.5%; Score 279.5; DB 5; Length 420;

Best Local Similarity 24.5%; Pred. No. 1.2e-12;

Matches 79; Conservative 62; Mismatches 166; Indels 15; Gaps 7;

Db 80 PAKSIIIGGQFAIWERGPPQERSRLCSI-ALSGMLGCFATALLIGFISLTGMPVFY 138

Y 139 IFGGVGCCLLMPFVYIDDPVSPWISTSEKEYIISLKGQVSSK--QPLPIKAMRS 196

Db 5 PTNNAITG-----NMFPSSSEKSTALSTFTLGNQIASAGSPMAAVACASLDGMPATFY 57

Y 139 IFGGVGCCLLMPFVYIDDPVSPWISTSEKEYIISLKGQVSSK--QPLPIKAMRS 196

Db 58 FAGIFATGWSLMPFTTASHPAKVMKMKKEKYLAVNVKHHSEKSTRSPYSKILTS 117

Y 197 LPISICLGC-FSHOWLSTWVVIPIYISSVYHVNINDNGLSLAPRTVAMVIGVGY 255

Db 118 -PAFLGQLQCHFEFVLFMTLQIYLPYFKEVHLGVIANGTFEFAIPNIFNMIFKVVWGI 176

Y 256 LADFLTKKFKFLIT-VKRIATILGSLPSALIVSLPIYNSGITRTALLTSCGLSTL-- 313

Db 177 GIDKIKKKNKILSNKAVVSHGVASFSSFLILALFPVDCSNPTTGLIFCLAMSSMGT 236

Y 313 CQSGIYINVLDIAPRYSFSLGASRGFSSIAPIVPTVSGFLLSODPFGMRNVFELFA 372

Db 237 FVSGGYTSLSLAPQYITMNAISMFVAMIGRLITTPAVMSFRKDKGTAAEMONITGCSL 296

Y 373 VNLGLFLYLIIFGADVQEMAK 394

Db 297 AHIFSGSIFLFGSGELQDMAK 318

RESULT 14

ID 093599 PRELIMINARY; PRT; 592 AA.

AC 093599;

DT 01-FEB-1997 (TREMBLrel. 02, Created)

DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)

DT 01-JAN-1999 (TREMBLrel. 09, Last annotation update)

DE T28F3.4 PROTEIN.

GN T28F3.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RN [1]

RP SEQUENCE FROM N.A.

RA MORTIMORE B.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

DR EMBL; Z82285; CAB05296.1; -

DR PFAM; PF00083; sugar-tri.1.

SO SEQUENCE 592 AA; 65489 MW; 46C99E3B CRC32;

Query Match 13.5%; Score 279.5; DB 5; Length 592;

Best Local Similarity 22.5%; Pred. No. 1.8e-12;

Matches 81; Conservative 74; Mismatches 146; Indels 59; Gaps 10;

Y 79 LPKSIILGGQFAIWERGPPQERSRLCSIALSG-MLGCFATALLIGFISLTGMPVF 137

Db 190 LPANNAIIA-----NMFPSEKSTALSTFTTGNQMGAGGNPVAASICASSFQPSIF 242

Y 138 YIFGGVGCCLLMPFVYIDDPVSPWISTSEKEYIISLKGQVSSKQPL-----PIK 191

Db 243 YFASIVSTIWSVCFLTASNOPSKCKVMKTERDYLDANVARSNKTNVSILRSLVPPS 302

Y 192 AMLRSLPTWSICLQCFSHOWLSTWVVIPIYISSVYHVNINDNGLSLAPRTVAMVIGM 251

Db 303 KILKSPAFLLAQLOCFVINFITLQIYLPNFKVHLGVIANGTYTSVPIVFAIKI 362

Y 252 VGGYIADFLIYK-----FLITVKKIATILGSLPSALIVSLPYL-----NSGYTAT 300

Db 363 VGIITIDAKKERRKIISGTFGVLSOSIANFGAL-----FLITITFYVDCNTPTLGFV--- 416

Y 301 ALLTISGISTL-----CQ-----SGIYINVLDIAPRYSFSLGASGFSSI 342

Db 416 -----FFCMASAIHSAHPINENSRTGCMGTIVSGYTSLSLAPRTVYTMSSIVFCAM 471

Y 343 APVIPTVSGFLLSODPFGMRNVFLLFAVNLGLFLYLIIFGADVQEM-----AKERK 397

Db 472 GRLATPATVGLIKKQNTISEMOTLFLVCAANIIIGAVFLVVGSGELQDMGLEDKAKMK 531

RESULT 15

ID 023063 PRELIMINARY; PRT; 478 AA.

AC 023063;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)

DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)

DE SIMILAR TO SODIUM-DEPENDENT PHOSPHATE COTRANSPORTERS.

GN T28H11.8.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;

RN Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.

RP [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RX MEDLINE; 94150718.

RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M.,

RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DUREIN R., FAVELLO A., FULTON L.,

RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,

RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,

RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,

RA PARSONS J., PERCY C., KIFKEN L., ROOPRA A., SANDERS D., SHOWNKEEN R.,

RA SMALDON N., SMITH A., SONNHAMER E., STADEN R., SOLSTON J.,

RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,

RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMANN P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans.";

RL Nature 368:32-38(1994).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN-BRISTOL N2;

RA NELSON J., WOHLDMANN P.;

RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.

RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U64609; AAB04605.1; -;
 SQ SEQUENCE 478 AA; 54373 MW; 0BE6E4A0 CRC32;

Query Match 12.5% Score 259; DB 5; Length 478;
 Best Local Similarity 26.4%; Pred. No. 4e-11;
 Matches 86; Conservative 51; Mismatches 153; Indels 36; Gaps 10;

QY	95	RMGPPQERSRLCSIALSGMLGCEFTAILIGFI--SETLGMPFVFIYIFGVGCYCLLMF	152
DB	138	RMTESESEASPFESIMLATSGFGPLEFTMILGEMCSSPFGWEATYIL-GVGTPISSYAF	196
QY	153	VVIYDDPVSYFW-ISTSEKEYIISLQOQVSSKOPFIKAMLSLPIWISICLCFSHOW	211
DB	197	AYYSDNVEKNKNLESESEKKYILAG--KHSYREKVPYKALKDSTIW-ISLMFTGY	253
QY	212	LVTSMVY--YIPYISSVYHVNIRDNGLSALPEIVAMVIGMVGYLADFLTKKRLI	268
DB	254	L--AMIVYQOISPTFIKQVLFHTRETFGFSALPOLAIFTKIGGRLLDFM--IFFL	308
QY	269	TVRKIAATILGSLPSSALIVSLPYLNS-----GYTATALLTSCGLST	311
DB	309	HVKECYKFGGPKLTLVPLLIIESMSAFSLFLTGFLDDRWSLIFMMIFASLHFFVPV	368
QY	312	LCOSGIYINVIDIAPRYSSTLMGASRGFSSIAPIVPTVSGFLLSQDPEFGWRNVFFLLF	371
DB	369	ICSRITQIR--AGOSHHEFALNLMVYAGIAQILIPLGVOAAVPEPTRSOWSFVFYPLV	424
QY	372	AVNLGLLEFYLLFGEADYQEWAKERK	397
DB	425	ITVVITSLITYLFSRAPPAEWTMKRR	450

Search completed: May 25, 2000, 14:51:46
 Job time: 2442 sec


```
337 rgglypseriserllealaprovalprothvalserglyp 353
|||||.....:|||||.....:|||||
294 GAGGATTGGACATTCATCGCTGCTGCTACCATTTGCTGCTT 245
354 LeuleSerGlnAspProgluPheglyTParGAsnValPhepheleule 370
|||||.....:|||||.....:|||||
244 TTCCATGAGCAGACTCTGAAATTCGGGTGGAGAAATTTCTTTTGTAGT 195
370 uphealavalasnleuleuglyleuleuetheryleullephglyGlna 387
|||||.....:|||||.....:|||||
194 GTTGGAGTTAACTACTGCTTAATCATCTACCTGCTTTGGGAAAG 145
387 laaspvalglnluTPalalaGlnaGlylsleuThargleu 401
|||||.....:|||||.....:|||||
144 CGATGTTCAAGATGGCTGAGAGAGAACTCACTCGTTTG 101

seq_name: gb_est12:AA276173

seq_documentation_block:
LOCUS AA276173 495 bp mRNA EST 01-APR-1997
DEFINITION vc31a07.t1 Barstead MPLRBI Mus musculus cDNA clone IMAGE:776148 5'
similar to TR:G887523 G887523 SODIUM-PHOSPHATE TRANSPORT SYSTEM 1.
; mRNA sequence.
ACCESSION AA276173
VERSION AA276173.1 GI:1918811
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 495)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The Washu-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1292357.
Contact: Marra M/Mouse EST Project
Washu-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.lbnl.gov) for further information.
MGI:469004
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 459.
Location/Qualifiers
1..495
/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/clone="IMAGE:776148"
/clone_11b="Barstead MPLRBI"
/sex="mixed"
/tissue_type="Kidney"
/dev_stage="6 weeks"
/lab_host="DH10B"
/note="Vector: pUT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: EcoRI; Site 2: NotI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCGAAATCGAAGGAGGCGGCGGCTTTTCTTTTCTTTTCTTTTCTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
[CATGATTCGCTAC], digested with Not I and cloned into the
Not I and Eco RI sites of the modified pUT73 vector.
Library constructed by Bob Barstead."

BASE COUNT 124 a 129 c 88 g 154 t
ORIGIN
```

```
alignment_scores:
Quality: 575.00 Length: 164
Ratio: 3.859 Gaps: 0
Percent Similarity: 90.854 Percent Identity: 67.073

alignment_block:
US-09-391-958-1 x AA276173 ..

Align seg 1/1 to: AA276173 from: 1 to: 495

157 AspAspProValSerTyrrProTrrPleSerThrsSerLutysgluTyrril 173
|||||.....:|||||.....:|||||
3 GATGACCCCTGCTCTCTACCCATGATNAGTGCCCAAGAGAGATATAT 52
|||||.....:|||||.....:|||||
173 eileSerSerleuLysGlnGlnValGlySerSerlyGlnProleuProI 190
|||||.....:|||||.....:|||||
53 TTTATCCTCTCTGACACACAGATTACAGCTCAGAAAGACAGCACTTCCA 102
|||||.....:|||||.....:|||||
190 lelysalamelleuargSerleuProileTrrPserileCysleuGlyCys 206
|||||.....:|||||.....:|||||
103 TCAAGCTATGCTCAATCTCTGCTCTGCTGCTCAGTGCTCTGCTGCACT 152
|||||.....:|||||.....:|||||
207 PheSerHisGlnTrpLeuValSerThrMetValValTyrrileProThry 223
|||||.....:|||||.....:|||||
153 ATGACCCATCATGCTGGCTTTAACACCTTTATATATGATACACTCCACTA 202
|||||.....:|||||.....:|||||
223 rilesSerValTyrrHisValasnileArgaspasnGlyLeuLeuSera 240
|||||.....:|||||.....:|||||
203 CACACAGTTCTGTAATCAAGTTAATCATCAGACAGCAATGGGTCTGCTT 252
|||||.....:|||||.....:|||||
240 laleupropheillevalaIaTrrValilleGlymetValGlyTyrrleu 256
|||||.....:|||||.....:|||||
253 CTCTCCCTTTATATGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
|||||.....:|||||.....:|||||
257 AlaSpPheleuThrlTyrrLysPheArgleuThleThrValArgLysil 273
|||||.....:|||||.....:|||||
303 GCAGATTTCTACTGACAGCAAGATTTTAACTGCTTAACCTGAGAAATTT 352
|||||.....:|||||.....:|||||
273 ealathilleuenglySerleuProSerSerAlaLeuileValSerleup 290
|||||.....:|||||.....:|||||
353 CACACAGCTTTTGAAGAAATGCTCTCTGCGACGCCCTGTTGGGCTCTGC 401
|||||.....:|||||.....:|||||
290 roTyrrleuasnSerGlyTyrrileThralaThralaleuThleuSer 306
|||||.....:|||||.....:|||||
402 C.TATATCCAAATCCAGCTATATTTACAAATTAATATTTCTGACAAATTC 451
|||||.....:|||||.....:|||||
307 CysGlyLeuSerThrlLeuCysGlnSerGlyTyrrileThleasn 320
|||||.....:|||||.....:|||||
452 TGTGACTGTGCCCTCTATCTCAGAGCAGAACTATATTAAT 493
|||||.....:|||||.....:|||||

seq_name: gb_est27:A1427938

seq_documentation_block:
LOCUS A1427938 452 bp mRNA EST 09-MAR-1999
DEFINITION mm25d12.x1 Stratagene mouse skin (#937313) Mus musculus cDNA clone
IMAGE:52251 3' similar to TR:000476 000476 SODIUM PHOSPHATE
TRANSPORTER. ; mRNA sequence.
ACCESSION A1427938
VERSION A1427938.1 GI:4273864
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 452)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Page,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shln,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.
The Washu-HMI Mouse EST Project 1999

TITLE The Washu-HMI Mouse EST Project 1999
```

JOURNAL Unpublished (1999)
On Feb 17, 1998 this sequence version replaced g1:2889559.
COMMENT Contact: Marra M/Mashu-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@wustl.edu
This clone is available royalty-free through LML; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new
data is from the 3' end
Possible reversed clone: similarity on wrong strand
High quality sequence stop: 400.

FEATURES

source
1..452
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:522551"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site: 1:
Ecort; Site: 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT 113 a 119 c 79 g 140 t 1 others

ORIGIN

alignment_scores:

Quality: 546.00 Length: 149
Ratio: 3.985 Gaps: 0
Percent Similarity: 91.946 Percent Identity: 66.443

alignment_block:

US-09-391-958-1 x A1427938 ..

Align seg 1/1 to: A1427938 from: 1 to: 452

```

158 ASPProVALSerTYrProTPILeserThSerGLuYSLuYrILleI 174
|||||
6 GACCTGTCTCTCAACCAAGTGAAGTGGCCCAAGAAAGAAATATATTTT 55
|||||
174 eSerSerLeuYsgInGlySerSerLeuYsgInProLeuProIleT 191
|||||
56 ATCCCTCTGAACCAACAGTTCAGCTCAGAGACAGCACTTCCATCA 105
|||||
191 ySAlaMetLeuArgSerLeuProIleTPSerILeCYSLeugLYCysAPhe 207
|||||
106 AAGCATCTCAAAATCTGCTCTGCTGCATGCTCTGCACTATG 155
|||||
208 SerHISGInTPLeuValSerThMetValValTYrILeProThTYrIL 224
|||||
156 ACCCAATCAGTGGCTGTACACCTTATATATGACACTCCACACTACAT 205
|||||
224 eSerSerValTYrHISValASnILeArgASpASnGlyLeuLeuSerAlaL 241
|||||
206 CAGTCTGTATTCAAAGTTAATCATCAGACAGCAATGGTCTCTCTCTC 255
|||||
241 euPProPheLeValAlaTPValILeGlyMetValGlyLYrTYrLeuLa 257
|||||
256 TTCCTTTATTTGTGCTGGGCTCTGTGATCTGGAGAGCTGGCTGCA 305
|||||
258 ASPPheLeuLeuThLYrSLysPheArgLeuLeuILeThrValArgLYsILeAl 274
|||||
306 GATTTTCTACTGACGAAGATTTTAGGCTCATACTGTGAGAAATTCAT 355

```

274 eThrILeugLYSerLeuProSerSerAlaLeuILeValSerLeuProT 291
|||||
356 CACACTTTTAGGAATGCTCTCTCTGACGCCCTGTTGGCGCTCTCCCT 405
|||||
291 yrLeuASnSergLYrILeThAlaThrAlaLeuLeuThLeuSer 306
|||||
406 ATATCCATCCAGCTATATACAAATATATATTTCTGCAAAATTCN 452
|||||

seq_name: gb:est9:AA073942

seq_documentation_block: 455 bp

LOCUS

AA073942

mm97h1.r1 Stratagene mouse heart (#937316) Mus musculus cDNA clone

IMAGE:536421 5' similar to TR:G165690 G165690 RENCAL CORTICAL

NA/P-I-COTRANSFORMER.; mRNA sequence.

ACCESSION

AA073942

AA073942.1 GI:1595689

VERSION

EST.

KEYWORDS

house mouse.

SOURCE

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE

1 (bases 1 to 455)

Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

AUTHORS

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

THEISING,B., WYLLIE,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

THE WASHU-HMI Mouse EST Project

Unpublished (1996)

On Sep 12, 1996 this sequence version replaced g1:1400896.

COMMENT

Contact: Marra M/Mouse EST Project

WASHU-HMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@wustl.edu

This clone is available royalty-free through LML; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:323357

Seg primer: -28n13 rev1 ET from Amersham

High quality sequence stop: 447.

Location/Qualifiers

1..453

/organism="Mus musculus"

source

/strain="NIH/Swiss"

/db_xref="taxon:10090"

/clone="IMAGE:536421"

/clone_lib="Stratagene mouse heart (#937316)"

/sex="pooled"

/tissue_type="heart"

/dev_stage="13 day embryos"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: heart; Vector: pBluescript SK-; Site: 1:

Ecort; Site: 2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. 93 pooled NIH/Swiss 13 day embryo hearts.

Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'

adaptor sequence: 5' GAATTCGCGACGAG 3' -3' adaptor

sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"

BASE COUNT

105 a 98 c 107 g 145 t

ORIGIN

alignment_scores:

Quality: 525.00 Length: 151
Ratio: 3.804 Gaps: 0
Percent Similarity: 91.391 Percent Identity: 66.887

alignment_block:

US-09-391-958-1 x AA073942 ..

Align seg 1/1 to: AA073942 from: 1 to: 455

seq_name: gb_est37:AI957130

alignment_block:

ation can

TITLE The Washu-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394568.
Contact: Marra M/Mouse EST Project

Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu
This clone is available royalty-free through LINDA; contact the
IMAGE Consortium (info@image.lind.gov) for further information.
MGI:316399

Possible reversed clone: similarity on wrong strand
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 314.

FEATURES

SOURCE

1. 438
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:522551"
/clone_lib="Stratagene mouse skin (#937313)"
/sex="females"
/tissue_type="whole skin"
/dev_stage="11 weeks old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: skin; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Whole skin from 11 week old C57BL/6 female mice.
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'
adaptor sequence: 5' GATTTCGCGACGAG 3' -3' adaptor
sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3' "

BASE COUNT 140 a 94 c 98 g 106 t
ORIGIN

alignment_scores:

Quality: 424.00 Length: 113
Ratio: 4.198 Gaps: 0
Percent Similarity: 89.381 Percent Identity: 69.912

alignment_block:
US-09-391-958-1 x AA068000/rev ..

Align seg 1/1 to reverse of: AA068000 from: 1 to: 438

```

289 leuprotyleuansserglytyrilethralathralaleuThyle 305
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
438 CTCGCCATATCCATCCACATCATATACACATTTATCTCGACAT 389
305 usercysglyleuserthleucysglylethylleasval 322
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
388 TTCCTGTGAGCTGTGCCCTATCTACGACGAGCAATCTAATATTCGT 339
322 euaspilaleaprogtyrserserpheleumetglyalaserargly 338
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
338 TAACATTTGCTCCAGATGATGCCAGCTTCTCATGGAGCAATCAAGAGA 289
339 pheserterlealeaprovalillevaliprothralvalserglypheleu 355
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
288 TTGGCAATTCATCGCGCTGTGCTGTACCATTTGCTGCTTTTCTTCT 239
355 userglaspproglyubheglytrpargasnvalphetheleuleubhea 372
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
238 CACCCAGAGACTCTGAATTCGCTGAGGAATTTCTTTTGTAGTGTG 189
372 lavalasleuleuglyleuetherthyleuilepheglyglualaspp 388
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
188 CAGTTAACCTACTGCTTATCATCTACCTCGTTTGGGAAAGACAGAT 139
389 valglnlutrpalatylsgluargylsleuthrargleu 401
|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

138 GTTCAGAAATGGCTAGAGAGAGAACCTCCTGTTG 100
seq_name: gp_est18:AA675103

seq_documentation_block: 462 bp mRNA EST 28-NOV-1997
LOCUS AA675103
DEFINITION vq03e06.r1 stratagene mouse diaphragm (#937303) Mus musculus cDNA
clone IMAGE:1093186 5' similar to TR:G450532 G450532
SODIUM-PHOSPHATE TRANSPORT SYSTEM 1. ; mRNA sequence.

ACCESSION AA675103
VERSION AA675103.1 GI:2652340

KEYWORDS
SOURCE

house mouse.
Mus musculus

REFERENCE
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Gelsel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lemon,G., Soares,B., Wilson,R. and
Waterston,R.
The Washu-HHMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1405024.
Contact: Marra M/Mouse EST Project

Washu-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.wustl.edu

This clone is available royalty-free through LINDA; contact the
IMAGE Consortium (info@image.lind.gov) for further information.
MGI:599418
Seq primer: -28m13 rev1 ET from Amersham
High quality sequence stop: 433.

FEATURES

SOURCE

1. 462
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1093186"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dT. Average
insert size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GATTTCGCGACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTCTTTTCTTTTCTTTT 3' "

BASE COUNT 102 a 120 c 108 g 131 t 1 others
ORIGIN

alignment_scores:

Quality: 416.00 Length: 179
Ratio: 3.382 Gaps: 2
Percent Similarity: 68.715 Percent Identity: 48.045

alignment_block:
US-09-391-958-1 x AA675103 ..

Align seg 1/1 to: AA675103 from: 1 to: 462

```

96 TrpglypropogingluargserarglyserlealealeuSergl 112
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2 TTGGCTCCGCACTTGAGACGAGCAAGCTCACACATGCGTGTCA.. 49
112 ymeluenuuglycyspherhralaleuileglyglypheileserg 129
|||||:|||||:|||||:|||||:|||||:|||||:|||||:
49 ..... 49

```

```

129 IuThIeuclYTrProPheValPheTYrIlePheGlyGlyValGlyCys 145
    |||:|||||
50 .....GTAAGCATTTGGCTGT 64
146 ValCysCysLeuLeuTrpPheValIleTYrAspAspProValSerTY 162
    |||:|||||
65 GTCTGCTGTCTGTGGTTCACGGGTGATTTATGATGACCCCATGATCA 114
162 rProTPPIleSerThrSerGluIysGluIuTYrIleSerSerLeuIysG 179
    :||| |||:|||||
115 CCCATGCAATAGTGTGAGGAAAAGAACATCATCTTCAGTGGCTC 164
179 InGIuValGlySerSerIysGluProLeuProIleIysAlaMetLeuArg 195
    |||:|||||
165 AGCAGTGTGAGTCTCCACAGACGCTGTCTCCATTAAGGAGTGTCA 214
196 SerLeuProIleTrpSerIleCysLeuGlyCysPheSerHisGlnTrpLe 212
    |||:|||||
215 TGCTGGCCACTGTGGCCATTTTCATGGGTTTTCAGTCATTTCTGGCT 264
212 uValSerThrMetValValTYrIleProThTYrIleSerSerValTYrH 229
    | :|:|||||
265 TTG.ACCTATATCATTAACATACCTACGACGTACATCAGCACAGTGTCTG 313
229 IsValAsnIleArgAspAsnGlyLeuLeuSerAlaLeuProPheIleVal 245
    |||:|||||
314 ACGTTAATCATCAGAGACGAGTGGGTCTGTGTCTCTCTCTCTCTCTCTCT 363
246 AlaTrpValIleGlyMetValGlyTYrLeuAlaAspPheLeuLeuTh 262
    |||:|||||
364 GCCTCAAGCTGTACAAATTTAGAGAGTCAGATGGCAGATTTCCATTCCTC 413
262 rLysLys...PheArgLeuIleThrValArgLysIle 273
    :|:|||||
414 CAGGAATCTTCTCAGCTTAATCATCAGTTCTGANAATCTC 450

```